Precoder Design for Signal Superposition in MIMO-NOMA Multicell Networks

Van-Dinh Nguyen, *Student Member, IEEE,* Hoang Duong Tuan, Trung Q. Duong, *Senior Member, IEEE,* H. Vincent Poor, *Fellow, IEEE,* and Oh-Soon Shin, *Member, IEEE*

***Abstract*—The throughput of users with poor channel conditions, such as those at a cell edge, is a bottleneck in wireless systems. A major part of the power budget must be allocated to serve these users in guaranteeing their quality-of-service (QoS) requirement, hampering QoS for other users and thus compromising the system reliability. In nonorthogonal multiple access (NOMA), the message intended for a user with a poor channel condition is decoded by itself and by another user with a better channel condition. The message intended for the latter is then successively decoded by itself after canceling the interference of the former. The overall information throughput is thus improved by this particular successive decoding and interference cancellation. This paper aims to design linear precoders/beamformers for signal superposition at the base stations of NOMA multi-input multi-output multi-cellular systems to maximize the overall sum throughput subject to the users' QoS requirements, which are imposed independently on the users' channel condition. This design problem is formulated as the maximization of a highly nonlinear and nonsmooth function subject to nonconvex constraints, which is very computationally challenging. Path-following algorithms for its solution, which invoke only a simple convex problem of moderate dimension at each iteration are developed. Generating a sequence of improved points, these algorithms converge at least to a local optimum. Extensive numerical simulations are then provided to demonstrate their merit.**

***Index Terms*—Multi-user interference system, multi-input multi-output (MIMO), nonorthogonal multiple access (NOMA), nonconvex optimization, quality-of-service (QoS), successive interference cancellation (SIC), signal superposition.**

This work was supported in part by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (Development of Beyond 5G Mobile Communication Technologies (Ultra-Reliable, Low-Latency, and Massive Connectivity) and Combined Access Technologies for Cellular-based Industrial Automation Systems), in part by the Australian Research Council's Discovery Projects under Project DP130104617, in part by the U.K. Royal Academy of Engineering Research Fellowship under Grant RF1415/14/22 and by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/P019374/1, and in part by the U.S. National Science Foundation under Grants CCF-1420575 and CNS-1456793. Part of this work has been submitted to IEEE Global Communication Conference (GLOBECOM 2017), Singapore, Dec. 2017.

V.-D. Nguyen and O.-S. Shin are with the School of Electronic Engineering and the Department of ICMC Convergence Technology, Soongsil University, Seoul 06978, Korea (e-mail: {nguyenvandinh, oshin}@ssu.ac.kr).

H. D. Tuan is with the Faculty of Engineering and Information Technology, University of Technology Sydney, Broadway, NSW 2007, Australia (email: tuan.hoang@uts.edu.au).

T. Q. Duong is with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast BT7 1NN, United Kingdom (e-mail: trung.q.duong@qub.ac.uk).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

I. Introduction

The explosive growth of mobile traffic demand is the driving force behind the development of signal processing and communication technologies to significantly upgrade the high-end experiences of communication such as high throughput, high reliability, and ubiquitous access. It is widely believed that interference-limited techniques such as coordinated multipoint transmission (CoMP) [1], [2], which treat interference as noise cannot meet the edge throughput requirements for 5G [3]. Dirty-paper coding (DPC) [4], under which the interference is successively mitigated, can improve both the edge and sum throughput but is difficult to implement in practice and remains only as a theoretical concept due to its high computational complexity. Nonorthogonal multiple access (NOMA) has been recently recognized as an essential enabling technology for 5G systems [5] due to its potential to improve the edge throughput [6].

In NOMA, a base station (BS) transmits a signal superposition to all users. The users are paired so that in each pair there is one with a better channel condition and another with a poorer channel condition [7]. The messages intended for each pair of users are sequentially decoded as follows. First, the message for the user with the poorer channel condition is decoded by both users. The message for the user with the better channel condition is then successively decoded by this user after canceling the interference from the other user [8]. Thus, while the throughput at the users with poorer channel condition remains the same as that in interference-limited techniques, the throughput at the users with better channel condition is clearly improved, leading to a higher system throughput.

Multi-input multi-output (MIMO) is widely known for its enormous potential in improving the capacity of wireless communication systems without requiring extra bandwidth or power. NOMA for MIMO communication (MIMO-NOMA) in single-cell systems for achieving higher throughput has been investigated in [9], [10], and an extension to multi-cell cases has been considered in [11]. In multi-cell systems, the effects of inter-cell interference are acute and unpredictable, limiting the quality-of-service

(QoS) for cell-edge users. It is therefore challenging to realize the benefit that NOMA may bring to multi-cell systems.

A. Related Works

In this subsection we discuss the state-of-the-art of signal processing techniques for NOMA downlink transmission. NOMA was mostly studied for single-cell multi-input single-output (MISO) systems known as MISO-NOMA, where the multiple-antenna BS broadcasts signal superpositions to single-antenna users. Under the assumption on low QoS requirement for the near user (with a good channel condition) and high QoS requirement for the far user (with a poor channel condition) in a two-user MISO-NOMA, [12] proposed a heuristic computational procedure with neither convergence nor optimality guaranteed for the beamforming power minimization. Under similar users' QoS requirements in a $2K$-user MISO-NOMA, it used a particular zero-forcing beamformer to cancel the inter-pair interference, so the problem of $2K$-user MISO-NOMA beamforming is decomposed into K independent subproblems of two-user MISO-NOMA beamforming. A closed-form solution for minimization of beamforming power in two-user MISO-NOMA subject to natural users' QoS requirements was obtained in [13]–[15]. In [16], users performed successive interference cancellation (SIC) based on the channel gain differences. Its proposed algorithm for beamforming is of high computational complexity.

Regarding MIMO-NOMA, [9] and [10] derived the outage probability experienced by users in zero-forcing post-coding or signal alignment. Power allocation for achieving the ergodic capacity of two-user MIMO NOMA was considered in [17] and [18]. User-pairing to enhance the throughput of users of poor channel condition was proposed in [19]. [11] proposed two interference alignment based coordinated beamforming for a two-cell MIMO-NOMA, where the interference at all cell-center users and edge-center users is canceled.

B. Motivation and Contributions

The paper considers the problem of designing linear precoders/beamformers at the BSs for MIMO-NOMA multi-cell systems to maximize their sum throughput while simultaneously meeting the users' QoS requirements. In general, such a design problem is very complicated as the objective function is nonlinear and nonsmooth, and the QoS constraints are highly nonconvex, for which even finding a feasible point is already challenging. The main contributions of the paper are three-fold:

- For MIMO-NOMA, two path-following optimization algorithms are proposed for computation, which at least converge to a locally optimal solution. At each iteration, the first algorithm invokes a convex quadratic program while the second algorithm invokes a semidefinite program (SDP). Both these convex programs are of moderate dimension so their computation is very efficient.
- Another path-following algorithm tailored for MISO-NOMA is proposed, which explores much simpler structures of the throughput functions in MISO systems for more efficient computation.
- The provided numerical results show the essential performance improvements of NOMA based systems compared to the conventional systems. The capability of NOMA to improve both edge and sum throughput is revealed.

C. Paper Organization and Notation

The rest of the paper is organized as follows. Section II presents the system model and formulates the problem. A convex quadratic programming based path-following algorithm for the MIMO-NOMA problem is developed in Section III, while another SDP based path-following algorithm is developed in Section IV. Section V devotes to computation for MISO-NOMA problem. Numerical results are provided in Section VI, and Section VII concludes the paper.

Notation. Bold-faced upper-case letters are used for matrices, bold-faced lower-case letters are used for vectors, and lower-case letters are used for for scalars. $\mathbf{I}_n$ is the identity matrix of size $n \times n$. $\mathbf{X}^H$, $\mathbf{X}^T$, and $\mathbf{X}^*$ are the Hermitian transpose, normal transpose, and conjugate of a matrix $\mathbf{X}$, respectively. The inner product $\langle \mathbf{X}, \mathbf{Y} \rangle$ of matrices $\mathbf{X}$ and $\mathbf{Y}$ is defined as trace$(\mathbf{X}^H\mathbf{Y})$. Denote by $\langle \mathbf{A} \rangle$ the trace of a matrix $\mathbf{A}$, and by $|\mathbf{A}|$ its determinant. $\|\cdot\|$ stands for matrix's Frobenius norm or vector's Euclidean norm. For Hermitian symmetric matrices $\mathbf{A} \succeq \mathbf{0}$ ($\mathbf{A} \succ \mathbf{0}$, resp.) means that $\mathbf{A}$ is a positive semidefinite (positive definite, resp.) matrix. Accordingly $\mathbf{A} \succeq \mathbf{B}$ ($\mathbf{A} \succ \mathbf{B}$, resp.) means $\mathbf{A} - \mathbf{B} \succeq \mathbf{0}$ ($\mathbf{A} - \mathbf{B} \succ \mathbf{0}$, resp.). $\mathbb{C}$ is the set of all complex numbers, and $\emptyset$ is an empty set. $\Re\{x\}$ denotes the real part of a complex number x. $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\eta}, \boldsymbol{Z})$ means that $\mathbf{x}$ is a random vector following a complex circular Gaussian distribution with mean $\boldsymbol{\eta}$ and covariance matrix $\boldsymbol{Z}$. $\nabla_{\mathbf{x}} f(\mathbf{x})$ is the gradient of function $f(\cdot)$ with respect to its variable $\mathbf{x}$. $\mathbb{E}\{\cdot\}$ denotes the expectation operator.

II. System Model and Problem Formulation

This section presents a system model for signal superposition in MIMO-NOMA multi-cell systems and formulates an optimization problem for the precoder design. Relations to CoMP and DPC are also briefly clarified.

A. Signal Processing Model

Consider a downlink system consisting of N cells, where the BS of each cell is equipped with N_t antennas to serve $2K$ users (UEs) within its cell as illustrated in Fig. 1. Each UE is equipped with N_r antennas. In each cell, there are K near UEs (cell-center UEs), which are located inside the

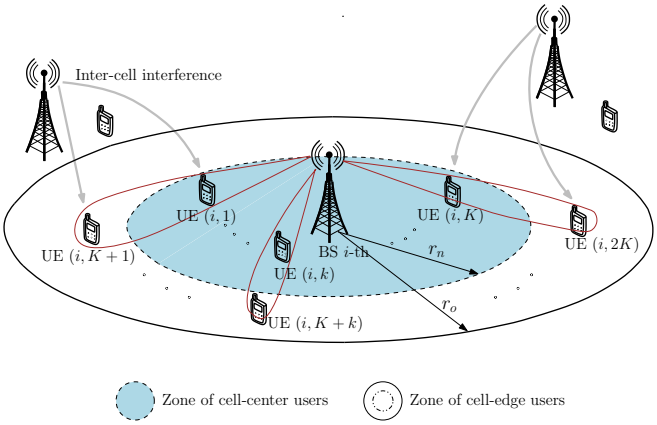


Fig. 1. An illustration of the cell of interest in a NOMA system serving $2K$ users.

circular area with radius r_n and the BS at its center, and K far UEs (cell-edge UEs), which are located within the ring area with inner radius r_n and outer radius r_o. K far UEs in each cell are not only in poorer channel conditions than other K near UEs but also are under more intensified inter-cell interference from adjacent cells.

Upon denoting $\mathcal{I} \triangleq \{1, 2, \cdots, N\}$ and $\mathcal{J} \triangleq \{1, 2, \cdots, 2K\}$, the j-th UE in the i-th cell is referred to as UE $(i, j) \in \mathcal{S} \triangleq \mathcal{I} \times \mathcal{J}$. The cell-center UEs are UE (i, j), $j \in \mathcal{K}_1 \triangleq \{1, \cdots, K\}$ while the cell-edge UEs are UE (i, j), $j \in \mathcal{K}_2 \triangleq \{K + 1, \cdots, 2K\}$. Thus the set of cell-center UEs and the set of cell-edge UEs are $\mathcal{S}_1 \triangleq \mathcal{I} \times \mathcal{K}_1$ and $\mathcal{S}_2 \triangleq \mathcal{I} \times \mathcal{K}_2$, respectively. In NOMA each cell-center UE $(i, j) \in \mathcal{S}_1$ is randomly paired with cell-edge UE $(i, p(j)) \in \mathcal{S}_2$ of the same cell to create a virtual cluster.[1] The signal superpositions are precoded at the BSs prior to being transmitted to the UEs. Specifically, the message intended for UE (i, j) is $\mathbf{s}_{i,j} \in \mathbb{C}^L$ with $\mathbb{E}\{\mathbf{s}_{i,j}(\mathbf{s}_{i,j})^H\} = \mathbf{I}_L$, which is precoded by matrix $\mathbf{V}_{i,j} \in \mathbb{C}^{N_t \times L}$, where L is the number of concurrent data streams and $L \leq \min\{N_t, N_r\}$. For notational convenience, let us define $\mathbf{V} \triangleq [\mathbf{V}_{i,j}]_{(i,j)\in\mathcal{S}}$. The received signals at UE (i, j) and UE $(i, p(j))$ are expressed as

$$\mathbf{y}_{i,j} = \mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\mathbf{s}_{i,j} + \mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)}\mathbf{s}_{i,p(j)} + \sum_{(s,l)\in\mathcal{S}\setminus\{(i,j),(i,p(j))\}} \mathbf{H}_{s,i,j}\mathbf{V}_{s,l}\mathbf{s}_{s,l} + \mathbf{n}_{i,j}, \quad (1)$$

and

$$\mathbf{y}_{i,p(j)} = \mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}\mathbf{s}_{i,p(j)} + \mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,j}\mathbf{s}_{i,j} + \sum_{(s,l)\in\mathcal{S}\setminus\{(i,j),(i,p(j))\}} \mathbf{H}_{s,i,p(j)}\mathbf{V}_{s,l}\mathbf{s}_{s,l} + \mathbf{n}_{i,p(j)} \quad (2)$$

where $\mathbf{H}_{s,i,j} \in \mathbb{C}^{N_r \times N_t}$ is the MIMO channel from the BS $s \in \mathcal{I}$ to UE $(i, j) \in \mathcal{S}$. The entries of the additive noise $\mathbf{n}_{i,j} \in \mathbb{C}^{N_r}$ are independent and identically distributed (i.i.d.) noise samples with zero mean and variance σ^2. The covariances of $\mathbf{y}_{i,j}$ and $\mathbf{y}_{i,p(j)}$ are thus

$$\mathcal{M}_{i,j}(\mathbf{V}) = \sum_{(s,l)\in\mathcal{S}} \mathbf{H}_{s,i,j}\mathbf{V}_{s,l}\mathbf{V}_{s,l}^H\mathbf{H}_{s,i,j}^H + \sigma^2\mathbf{I}_{N_r}, \quad (3)$$

and

$$\mathcal{M}_{i,p(j)}(\mathbf{V}) = \sum_{(s,l)\in\mathcal{S}} \mathbf{H}_{s,i,p(j)}\mathbf{V}_{s,l}\mathbf{V}_{s,l}^H\mathbf{H}_{s,i,p(j)}^H + \sigma^2\mathbf{I}_{N_r}. \quad (4)$$

The purpose of the paper is to design complex-valued precoding matrices $\mathbf{V}_{i,j}$ to maximize the overall spectral efficiency under a given pairing for NOMA. The MIMO channel states $\mathbf{H}_{s,i,j}$ are assumed unchanged during message transmission but may change independently from one message to another and are perfectly known at all nodes [10], [12], [16].

B. Problem Formulation

In NOMA, the message $\mathbf{s}_{i,p(j)}$ intended for the cell-edge UE $(i, p(j))$ is decoded by both UE $(i, p(j))$ and UE (i, j). Then message $\mathbf{s}_{i,j}$ intended for the cell-center UE (i, j) is decoded by itself only.

The cell-edge UE $(i, p(j))$ decodes its own message $\mathbf{s}_{i,p(j)}$ with the achievable rate

$$\mathcal{R}_{i,p(j)}^{p(j)}(\mathbf{V}) = \ln\Big|\mathbf{I}_L + (\mathbf{V}_{i,p(j)})^H\mathbf{H}_{i,i,p(j)}^H\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V})^{-1}\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}\Big| \quad (5)$$

where $\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V})$ is defined by

$$\begin{aligned}\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V}) &\triangleq \mathcal{M}_{i,p(j)}(\mathbf{V}) - \mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}(\mathbf{V}_{i,p(j)})^H\mathbf{H}_{i,i,p(j)}^H \\ &= \sum_{(s,l)\in\mathcal{S}\setminus\{(i,p(j))\}} \mathbf{H}_{s,i,p(j)}\mathbf{V}_{s,l}\mathbf{V}_{s,l}^H\mathbf{H}_{s,i,p(j)}^H + \sigma^2\mathbf{I}_{N_r}. \quad (6)\end{aligned}$$

On the other hand, the cell-center UE (i, j) decodes the message $\mathbf{s}_{i,p(j)}$ with the achievable rate

$$\mathcal{R}_{i,p(j)}^{j}(\mathbf{V}) = \ln\Big|\mathbf{I}_L + (\mathbf{V}_{i,p(j)})^H\mathbf{H}_{i,i,j}^H\mathcal{M}_{i,j}^{p(j)}(\mathbf{V})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)}\Big| \quad (7)$$

where

$$\begin{aligned}\mathcal{M}_{i,j}^{p(j)}(\mathbf{V}) &\triangleq \mathcal{M}_{i,j}(\mathbf{V}) - \mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)}(\mathbf{V}_{i,p(j)})^H\mathbf{H}_{i,i,j}^H \\ &= \sum_{(s,l)\in\mathcal{S}\setminus\{(i,p(j))\}} \mathbf{H}_{s,i,j}\mathbf{V}_{s,l}\mathbf{V}_{s,l}^H\mathbf{H}_{s,i,j}^H + \sigma^2\mathbf{I}_{N_r}. \quad (8)\end{aligned}$$

Hence, the throughput by decoding the message $\mathbf{s}_{i,p(j)}$ by UEs $(i, p(j))$ and (i, j) is

$$\mathcal{R}_{i,p(j)} = \min\Big\{\mathcal{R}_{i,p(j)}^{j}(\mathbf{V}), \mathcal{R}_{i,p(j)}^{p(j)}(\mathbf{V})\Big\}. \quad (9)$$

[1] Using more sophisticated user-pairing strategies may improve the performance of MIMO-NOMA networks (see e.g. [8], [19]) but it is beyond the scope of this paper.

Next, the throughput by decoding the message $\mathbf{s}_{i,j}$ by the cell-center UE (i,j) after decoding the message $\mathbf{s}_{i,p(j)}$ is

$$\mathcal{R}_{i,j}(\mathbf{V}) = \ln\Big|\mathbf{I}_L + (\mathbf{V}_{i,j})^H\mathbf{H}_{i,i,j}^H\mathcal{M}_{i,j}^p(\mathbf{V})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\Big| \tag{10}$$

where

$$\mathcal{M}_{i,j}^p(\mathbf{V}) \triangleq \mathcal{M}_{i,j}^{p(j)}(\mathbf{V}) - \mathbf{H}_{i,i,j}\mathbf{V}_{i,j}(\mathbf{V}_{i,j})^H\mathbf{H}_{i,i,j}^H$$
$$= \sum_{(s,l)\in\mathcal{S}\setminus\{(i,p(j)),(i,j)\}} \mathbf{H}_{s,i,j}\mathbf{V}_{s,l}\mathbf{V}_{s,l}^H\mathbf{H}_{s,i,j}^H + \sigma^2\mathbf{I}_{N_r}. \tag{11}$$

Our goal is to maximize the total sum throughput of the system under QoS for each individual UE and power budget at each BS, which is mathematically formulated as

$$\underset{\mathbf{V}}{\text{maximize}}\ \mathcal{P}(\mathbf{V}) \triangleq \sum_{i=1}^{N}\sum_{j=1}^{K}\Big(\mathcal{R}_{i,j}(\mathbf{V}) + \mathcal{R}_{i,p(j)}(\mathbf{V})\Big) \tag{12a}$$

$$\text{subject to } \mathcal{R}_{i,j}(\mathbf{V}) \geq r_{i,j},\ \forall i \in \mathcal{I},\ \forall j \in \mathcal{K}_1, \tag{12b}$$

$$\mathcal{R}_{i,p(j)}(\mathbf{V}) \geq r_{i,p(j)},\ \forall i \in \mathcal{I},\ \forall j \in \mathcal{K}_2, \tag{12c}$$

$$\sum_{j\in\mathcal{J}}\Big\langle \mathbf{V}_{i,j}\mathbf{V}_{i,j}^H\Big\rangle \leq P_i^{\max},\ \forall i \in \mathcal{I} \tag{12d}$$

where $P_i^{\max}$ in (12d) is the transmit power budget of BS i. The QoS constraints (12c) and (12d) set a minimum throughput requirement $r_{i,j}$ at the UE (i,j) and $r_{i,p(j)}$ at the UE $(i,p(j))$.

C. Relations to CoMP and DPC

In CoMP [1], [2], the problem of maximizing the sum throughput under QoS constraints is formulated as [20]

$$\underset{\mathbf{V}}{\text{maximize}}\ \mathcal{P}^{\text{CoMP}}(\mathbf{V}) \triangleq \sum_{i=1}^{N}\sum_{j=1}^{2K}\mathcal{R}_{i,j}^{'}(\mathbf{V}) \tag{13a}$$

$$\text{subject to } \mathcal{R}_{i,j}^{'}(\mathbf{V}) \geq r_{i,j},\ \forall i \in \mathcal{I},\ \forall j \in \mathcal{J}, \tag{13b}$$

$$\sum_{j\in\mathcal{J}}\Big\langle \mathbf{V}_{i,j}\mathbf{V}_{i,j}^H\Big\rangle \leq P_i^{\max},\ \forall i \in \mathcal{I} \tag{13c}$$

where $\mathcal{R}_{i,j}^{'}(\mathbf{V})$ is given by

$$\mathcal{R}_{i,j}^{'}(\mathbf{V}) = \ln\Big|\mathbf{I}_L + (\mathbf{V}_{i,j})^H\mathbf{H}_{i,i,j}^H\mathcal{M}_{i,j}^j(\mathbf{V})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\Big| \tag{14}$$

with

$$\mathcal{M}_{i,j}^j(\mathbf{V}) \triangleq \sum_{(s,l)\in\mathcal{S}\setminus\{(i,j)\}} \mathbf{H}_{s,i,j}\mathbf{V}_{s,l}\mathbf{V}_{s,l}^H\mathbf{H}_{s,i,j}^H + \sigma^2\mathbf{I}_{N_r}. \tag{15}$$

Compared $\mathcal{R}_{i,j}^{'}(\mathbf{V})$ defined by (14) to $\mathcal{R}_{i,p(j)}(\mathbf{V})$ and $\mathcal{R}_{i,j}(\mathbf{V})$ defined by (9) and (10) one can see that

$$\mathcal{R}_{i,p(j)}^{'}(\mathbf{V}) = \mathcal{R}_{i,p(j)}^{p(j)}(\mathbf{V}) \tag{16}$$

$$\geq \min\Big\{\mathcal{R}_{i,p(j)}^{j}(\mathbf{V}), \mathcal{R}_{i,p(j)}^{p(j)}(\mathbf{V})\Big\} \tag{17}$$

$$= \mathcal{R}_{i,p(j)},\ \forall j \in \mathcal{K}_2, \tag{18}$$

and

$$\mathcal{R}_{i,j}^{'}(\mathbf{V}) \leq \mathcal{R}_{i,j}(\mathbf{V}),\ \forall j \in \mathcal{K}_1, \tag{19}$$

i.e., under the same precoder $\mathbf{V}$, the throughput at cell-edge UEs is higher with CoMP while that at cell-center UEs is higher with NOMA. Thus, under the same precoder, NOMA does not need to perform better than CoMP in terms of the total throughput. The preference of NOMA now critically depends on its performance at its optimal precoder, which is sought in the next sections.

On the other hand, DPC at the BS i with encoding order from UE $(i,2K)$ to UE $(i,1)$ enables UE (i,j) view the massages intended for UEs (i,j'), $j' > j$ as non-causually known and thus cancel them from its received signal. Hence, the throughput by decoding the message $\mathbf{s}_{i,j}$ for UE (i,j) is defined by

$$\mathcal{R}_{i,j}^{''}(\mathbf{V}) = \ln\Big|\mathbf{I}_L + (\mathbf{V}_{i,j})^H\mathbf{H}_{i,i,j}^H\mathcal{M}_{i,j}^i(\mathbf{V})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\Big| \tag{20}$$

where

$$\mathcal{M}_{i,j}^i(\mathbf{V}) \triangleq \sum_{s\neq i}^{N}\sum_{l=1}^{2K}\mathbf{H}_{s,i,j}\mathbf{V}_{s,l}\mathbf{V}_{s,l}^H\mathbf{H}_{s,i,j}^H + \sum_{k=1}^{j-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,k}\mathbf{V}_{i,k}^H\mathbf{H}_{i,i,j}^H + \sigma^2\mathbf{I}_{N_r}. \tag{21}$$

The problem of maximizing the sum throughput under QoS constraints is formulated as

$$\underset{\mathbf{V}}{\text{maximize}}\ \mathcal{P}^{\text{DPC}}(\mathbf{V}) \triangleq \sum_{i=1}^{N}\sum_{j=1}^{2K}\mathcal{R}_{i,j}^{''}(\mathbf{V}) \tag{22a}$$

$$\text{subject to } \mathcal{R}_{i,j}^{''}(\mathbf{V}) \geq r_{i,j},\ \forall i \in \mathcal{I},\ \forall j \in \mathcal{J}, \tag{22b}$$

$$\sum_{j\in\mathcal{J}}\Big\langle \mathbf{V}_{i,j}\mathbf{V}_{i,j}^H\Big\rangle \leq P_i^{\max},\ \forall i \in \mathcal{I}. \tag{22c}$$

Apparently, under DPC the intra-cell interference from UEs of better channel condition is canceled in decoding the messages intended for UEs of poorer channel condition resulting in much better edge throughput compared with CoMP. NOMA is advantageous over DPC in terms of ease of implementation. By increasing the QoSs' requirement in (12b), (13b) and (22b), we will see that the NOMA's sum throughput is fully superior over CoMP's one and catches up DPC's one.

III. Convex Quadratic Based Iterations

Note that the QoS constraints (12b) and (12c) in (12), and (13b) in (13) are set beforehand, which are dependent on the UEs' throughput requirements but independent on their channel condition. In maximizing the sum throughput objective (13a), most of CoMP techniques (see e.g. [21]) are unable to address the QoS constraints (13b). Without setting such QoS constraints, the UEs of poor channel condition are easily disconnected from the service because it is well known that almost all of transmit power will be allocated to a very few UEs of the best channel conditions in maximizing the sum throughput, causing almost zero throughput at other UEs. The weighted sum throughput

maximization is only an ad hoc way to balance the UEs' throughput. Alternatively, the throughput satisfaction can be effectively handled via maximizing the users' worst throughput but the latter involves optimization of a nonsmooth objective function, for which these techniques are powerless. Both QoS constrained sum throughput maximization problem (13) and UEs' worst throughput maximization problem could be addressed very recently in [20].

Although the optimization problem (12) is different from (13) and (22), all functions appearing in the former have a similar structure to that appearing in the latter. Therefore, a systematic approach to solve the latter is expected to be applicable to the former. In this section, we adopt the approach of [20] to address (12). Unlike [22], which aims at expressing the nonsmooth function $\mathcal{R}_{i,p(j)}(\mathbf{V})$ in (9) and then the nonsmooth objective function in (12b) as d.c. (difference of two convex functions) by using the universality of d.c. functions [23] and leads to d.c. iterations [24] of high computational complexity, we will see now that each below iteration invokes only a simple convex quadratic program of low computational complexity.

Suppose that $\mathbf{V}^{(\kappa)} \triangleq [\mathbf{V}_{i,j}^{(\kappa)}]_{(i,j)\in\mathcal{S}}$ is a feasible point found at the $(\kappa-1)$-th iteration. Define the following quadratic functions in $\mathbf{V}$:

$$\begin{aligned}
\mathcal{R}_{i,p(j)}^{j,(\kappa)}(\mathbf{V}) &\triangleq a_{i,p(j)}^{j,(\kappa)} + 2\Re\Big\{\big\langle \mathcal{A}_{i,p(j)}^{j,(\kappa)}, \mathbf{V}_{i,p(j)}\big\rangle\Big\} \\
&\quad - \Big\langle \mathcal{M}_{i,j}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1} - \mathcal{M}_{i,j}(\mathbf{V}^{(\kappa)})^{-1}, \mathcal{M}_{i,j}(\mathbf{V})\Big\rangle, \\
\mathcal{R}_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V}) &\triangleq a_{i,p(j)}^{p(j),(\kappa)} + 2\Re\Big\{\big\langle \mathcal{A}_{i,p(j)}^{p(j),(\kappa)}, \mathbf{V}_{i,p(j)}\big\rangle\Big\} \\
&\quad - \Big\langle \mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1} - \mathcal{M}_{i,p(j)}(\mathbf{V}^{(\kappa)})^{-1}, \mathcal{M}_{i,p(j)}(\mathbf{V})\Big\rangle, \\
\mathcal{R}_{i,j}^{(\kappa)}(\mathbf{V}) &\triangleq a_{i,j}^{(\kappa)} + 2\Re\Big\{\big\langle \mathcal{A}_{i,j}^{(\kappa)}, \mathbf{V}_{i,j}\big\rangle\Big\} \\
&\quad - \Big\langle \mathcal{M}_{i,j}^{p}(\mathbf{V}^{(\kappa)})^{-1} - \mathcal{M}_{i,j}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1}, \mathcal{M}_{i,j}^{p(j)}(\mathbf{V})\Big\rangle \quad (23)
\end{aligned}$$

where $a_{i,p(j)}^{j,(\kappa)}$, $a_{i,p(j)}^{p(j),(\kappa)}$, $a_{i,j}^{(\kappa)}$ are given as

$$\begin{aligned}
a_{i,p(j)}^{j,(\kappa)} &\triangleq \mathcal{R}_{i,p(j)}^{j}(\mathbf{V}^{(\kappa)}) - \Big\langle (\mathbf{V}_{i,p(j)}^{(\kappa)})^H \mathbf{H}_{i,i,j}^H \\
&\qquad \times \mathcal{M}_{i,j}(\mathbf{V}^{(\kappa)})^{-1} \mathbf{H}_{i,i,j} \mathbf{V}_{i,p(j)}^{(\kappa)} \Big\rangle < 0, \\
a_{i,p(j)}^{p(j),(\kappa)} &\triangleq \mathcal{R}_{i,p(j)}^{p(j)}(\mathbf{V}^{(\kappa)}) - \Big\langle (\mathbf{V}_{i,p(j)}^{(\kappa)})^H \mathbf{H}_{i,i,p(j)}^H \\
&\qquad \times \mathcal{M}_{i,p(j)}(\mathbf{V}^{(\kappa)})^{-1} \mathbf{H}_{i,i,p(j)} \mathbf{V}_{i,p(j)}^{(\kappa)} \Big\rangle < 0, \\
a_{i,j}^{(\kappa)} &\triangleq \mathcal{R}_{i,j}(\mathbf{V}^{(\kappa)}) - \Big\langle (\mathbf{V}_{i,j}^{(\kappa)})^H \mathbf{H}_{i,i,j}^H \\
&\qquad \times \mathcal{M}_{i,j}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1} \mathbf{H}_{i,i,j} \mathbf{V}_{i,j}^{(\kappa)} \Big\rangle < 0, \qquad (24)
\end{aligned}$$

and $\mathcal{A}_{i,p(j)}^{j,(\kappa)}$, $\mathcal{A}_{i,p(j)}^{p(j),(\kappa)}$, $\mathcal{A}_{i,j}^{(\kappa)}$ are given as

$$\begin{aligned}
\mathcal{A}_{i,p(j)}^{j,(\kappa)} &\triangleq \mathbf{H}_{i,i,j}^H \mathcal{M}_{i,j}(\mathbf{V}^{(\kappa)})^{-1} \mathbf{H}_{i,i,j} \mathbf{V}_{i,p(j)}^{(\kappa)}, \\
\mathcal{A}_{i,p(j)}^{p(j),(\kappa)} &\triangleq \mathbf{H}_{i,i,p(j)}^H \mathcal{M}_{i,p(j)}(\mathbf{V}^{(\kappa)})^{-1} \mathbf{H}_{i,i,p(j)} \mathbf{V}_{i,p(j)}^{(\kappa)}, \\
\mathcal{A}_{i,j}^{(\kappa)} &\triangleq \mathbf{H}_{i,i,j}^H \mathcal{M}_{i,j}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1} \mathbf{H}_{i,i,j} \mathbf{V}_{i,j}^{(\kappa)}. \qquad (25)
\end{aligned}$$

Note that all functions in (23) are concave due to (6), (8), and (11):

$$\begin{aligned}
\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1} - \mathcal{M}_{i,p(j)}(\mathbf{V}^{(\kappa)})^{-1} &\succeq \mathbf{0}, \\
\mathcal{M}_{i,j}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1} - \mathcal{M}_{i,j}(\mathbf{V}^{(\kappa)})^{-1} &\succeq \mathbf{0}, \\
\mathcal{M}_{i,j}^{p}(\mathbf{V}^{(\kappa)})^{-1} - \mathcal{M}_{i,j}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1} &\succeq \mathbf{0}.
\end{aligned}$$

The following result shows that the complicated function defined by (9) and (10) is lower bounded by concave quadratic functions.

Theorem 1: For $\mathcal{R}_{i,p(j)}^{(\kappa)}(\mathbf{V}) \triangleq \min\Big\{\mathcal{R}_{i,p(j)}^{j,(\kappa)}(\mathbf{V}), \mathcal{R}_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V})\Big\}$ it is true that

$$\begin{aligned}
\mathcal{R}_{i,p(j)}(\mathbf{V}^{(\kappa)}) &= \mathcal{R}_{i,p(j)}^{(\kappa)}(\mathbf{V}^{(\kappa)}) \quad \text{and} \\
\mathcal{R}_{i,p(j)}(\mathbf{V}) &\geq \mathcal{R}_{i,p(j)}^{(\kappa)}(\mathbf{V}),\ \forall \mathbf{V}, \qquad (26)
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{R}_{i,j}(\mathbf{V}^{(\kappa)}) &= \mathcal{R}_{i,j}^{(\kappa)}(\mathbf{V}^{(\kappa)}) \quad \text{and} \\
\mathcal{R}_{i,j}(\mathbf{V}) &\geq \mathcal{R}_{i,j}^{(\kappa)}(\mathbf{V}),\ \forall \mathbf{V}. \qquad (27)
\end{aligned}$$

Proof: See [20, Appendix B]. ■

Based on these results, at the κ-th iteration, the following convex program, which is an inner approximation for the nonconvex optimization problem (12), is solved to generate the next feasible point $\mathbf{V}^{(\kappa+1)}$:

$$\begin{aligned}
\max_{\mathbf{V}} \quad & \mathcal{P}^{(\kappa)}(\mathbf{V}) \triangleq \sum_{i=1}^{N}\sum_{j=1}^{K}\Big(\mathcal{R}_{i,j}^{(\kappa)}(\mathbf{V}) + \mathcal{R}_{i,p(j)}^{(\kappa)}(\mathbf{V})\Big) && (28a) \\
\text{subject to} \quad & \mathcal{R}_{i,j}^{(\kappa)}(\mathbf{V}) \geq r_{i,j},\ \forall i \in \mathcal{I},\ \forall j \in \mathcal{K}_1, && (28b) \\
& \mathcal{R}_{i,p(j)}^{(\kappa)}(\mathbf{V}) \geq r_{i,p(j)},\ \forall i \in \mathcal{I},\ \forall j \in \mathcal{K}_2, && (28c) \\
& \text{(12d)}. && (28d)
\end{aligned}$$

A pseudo-code of this quadratic programing (QP)-based path-following procedure is given by Algorithm 1.

Note that $\mathbf{V}^{(\kappa)}$ is also feasible for (28) with $\mathcal{P}(\mathbf{V}^{(\kappa)}) = \mathcal{P}^{(\kappa)}(\mathbf{V}^{(\kappa)})$ by the equalities in (26) and (27). It is then true that $\mathcal{P}^{(\kappa)}(\mathbf{V}^{(\kappa+1)}) > \mathcal{P}^{(\kappa)}(\mathbf{V}^{(\kappa)}) = \mathcal{P}(\mathbf{V}^{(\kappa)})$ whenever $\mathbf{V}^{(\kappa+1)} \neq \mathbf{V}^{(\kappa)}$. Together with $\mathcal{P}(\mathbf{V}^{(\kappa+1)}) \geq \mathcal{P}^{(\kappa)}(\mathbf{V}^{(\kappa)})$ and according to the inequalities in (26) and (27), we have $\mathcal{P}(\mathbf{V}^{(\kappa+1)}) > \mathcal{P}(\mathbf{V}^{(\kappa)})$, i.e., the optimal solution $\mathbf{V}^{(\kappa+1)}$ of the convex quadratic problem (28) is a better point for the nonconvex nonsmooth optimization problem (12) than $\mathbf{V}^{(\kappa)}$. Therefore, once initialized from an feasible point $\mathbf{V}^{(0)}$), the sequence $\{\mathbf{V}^{(\kappa)}\}$ obtained by solving (28) is of improved feasible points for (12). By following the same arguments as those in [20, Proposition 1], we can prove that Algorithm 1 converges to a Karush-Kuh-Tucker (KKT) point of (12).

Generation of the initial points: A feasible point for constraints (12b)-(12d) for initializing Algorithm 1 is found

Algorithm 1 QP-based path-following algorithm for the STM (12) in MIMO-NOMA

Initialization: Set $\kappa := 0$ and solve (30) to generate an initial feasible point $\mathbf{V}^{(0)}$ for constraints (12b)-(12d).

1: **repeat**
2: Solve the convex quadratic program (28) to obtain the optimal solution: $\mathbf{V}^\star$.
3: Update $\mathbf{V}^{(\kappa+1)} := \mathbf{V}^\star$.
4: Set $\kappa := \kappa + 1$.
5: **until** Convergence

via the following problem of QoS feasibility

$$\max_{\mathbf{V}} \min_{(i,j)\in\mathcal{S}} \min\left\{\frac{\mathcal{R}_{i,j}(\mathbf{V})}{r_{i,j}}, \frac{\mathcal{R}_{i,p(j)}(\mathbf{V})}{r_{i,p(j)}}\right\} \quad (29a)$$

$$\text{subject to} \quad \text{(12d)}. \quad (29b)$$

Initialized from a feasible point $\mathbf{V}^{(0)}$ for the convex constraint (12d), the following iterations are invoked

$$\max_{\mathbf{V}} \min_{(i,j)\in\mathcal{S}} \min\left\{\frac{\mathcal{R}^{(\kappa)}_{i,j}(\mathbf{V})}{r_{i,j}}, \frac{\mathcal{R}^{(\kappa)}_{i,p(j)}(\mathbf{V})}{r_{i,p(j)}}\right\} \quad (30a)$$

$$\text{subject to} \quad \text{(12d)} \quad (30b)$$

till reaching a value more than or equal to 1 in satisfying (12b)-(12d).

Complexity Analysis: Problem (28) is convex quadratic with $m_{\text{QP}} = N(1+3K)$ quadratic constraints and $n = KN(2N_tL+1)$ real decision variables. Its computational complexity is $\mathcal{O}(n^2 m_{\text{QP}}^{2.5} + m_{\text{QP}}^{3.5})$.

IV. Semi-Definite Programming Based Iterations

To further improve the convergence speed of solving (12), we need explore more partial convex structures of functions (9) and (10). In this section, we propose a novel SDP-based path-following algorithm for (12). To this end, we will use the following matrix inequalities:

$$\begin{aligned}\mathbf{V}^H\mathbf{X}^{-1}\mathbf{V} \;\succeq\; & \bar{\boldsymbol{V}}^H\bar{\boldsymbol{X}}^{-1}\mathbf{V} + \mathbf{V}^H\bar{\boldsymbol{X}}^{-1}\bar{\boldsymbol{V}} \\ & - \bar{\boldsymbol{V}}^H\bar{\boldsymbol{X}}^{-1}\mathbf{X}\bar{\boldsymbol{X}}^{-1}\bar{\boldsymbol{V}}, \\ & \forall\ \mathbf{V},\bar{\boldsymbol{V}},\mathbf{X}\succ\mathbf{0},\bar{\boldsymbol{X}}\succ\mathbf{0}, \end{aligned} \quad (31)$$

and

$$\begin{aligned}\ln|\mathbf{X}| \;\geq\; & \ln|\bar{\boldsymbol{X}}| - \langle\bar{\boldsymbol{X}}, \mathbf{X}^{-1} - \bar{\boldsymbol{X}}^{-1}\rangle, \\ & \forall\ \mathbf{X}\succ\mathbf{0}, \bar{\boldsymbol{X}}\succ\mathbf{0}, \end{aligned} \quad (32)$$

whose proofs are given by Appendix A and Appendix B.

Let us treat the rate function $\mathcal{R}^j_{i,p(j)}(\mathbf{V})$ from (7) first. Applying (31) yields

$$\mathbf{V}^H_{i,p(j)}\mathbf{H}^H_{i,i,j}\mathcal{M}^{p(j)}_{i,j}(\mathbf{V})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)} \succeq \widetilde{\mathcal{Q}}^{j,(\kappa)}_{i,p(j)}(\mathbf{V}) \quad (33)$$

for

$$\begin{aligned}\widetilde{\mathcal{Q}}^{j,(\kappa)}_{i,p(j)}(\mathbf{V}) \triangleq\; & (\mathbf{V}^{(\kappa)}_{i,p(j)})^H\mathbf{H}^H_{i,i,j}\mathcal{M}^{p(j)}_{i,j}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)} \\ & + (\mathbf{V}_{i,p(j)})^H\mathbf{H}^H_{i,i,j}\mathcal{M}^{p(j)}_{i,j}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}^{(\kappa)}_{i,p(j)} \\ & - (\mathbf{V}^{(\kappa)}_{i,p(j)})^H\mathbf{H}^H_{i,i,j}\mathcal{M}^{p(j)}_{i,j}(\mathbf{V}^{(\kappa)})^{-1}\mathcal{M}^{p(j)}_{i,j}(\mathbf{V}) \\ & \times\mathcal{M}^{p(j)}_{i,j}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}^{(\kappa)}_{i,p(j)}, \end{aligned} \quad (34)$$

which also satisfies

$$\begin{aligned}&\widetilde{\mathcal{Q}}^{j,(\kappa)}_{i,p(j)}(\mathbf{V}^{(\kappa)}) = \\ &\quad (\mathbf{V}^{(\kappa)}_{i,p(j)})^H\mathbf{H}^H_{i,i,j}\mathcal{M}^{p(j)}_{i,j}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}^{(\kappa)}_{i,p(j)}. \end{aligned} \quad (35)$$

For $\mathcal{M}^{p(j)}_{i,j}(\mathbf{V})$ defined from (8), applying (31) again yields

$$\mathcal{M}^{p(j)}_{i,j}(\mathbf{V}) \;\succeq\; \mathcal{L}^{p(j),(\kappa)}_{i,j}(\mathbf{V}) + \sigma^2\mathbf{I}_{N_r} \quad (36)$$

over the trust region

$$\mathcal{L}^{p(j),(\kappa)}_{i,j}(\mathbf{V}) \succeq \mathbf{0}, \quad (37)$$

for the linear mapping

$$\begin{aligned}\mathcal{L}^{p(j),(\kappa)}_{i,j}(\mathbf{V}) \triangleq & \sum_{(s,l)\in\mathcal{S}\setminus\{(i,p(j))\}} \mathbf{H}_{s,i,j}\Big[\mathbf{V}_{s,l}(\mathbf{V}^{(\kappa)}_{s,l})^H \\ & + \mathbf{V}^{(\kappa)}_{s,l}\mathbf{V}^H_{s,l} - \mathbf{V}^{(\kappa)}_{s,l}(\mathbf{V}^{(\kappa)}_{s,l})^H\Big]\mathbf{H}^H_{s,i,j}. \end{aligned} \quad (38)$$

It follows from (33) that

$$\widetilde{\mathcal{Q}}^{j,(\kappa)}_{i,p(j)}(\mathbf{V}) \preceq \mathcal{Q}^{j,(\kappa)}_{i,p(j)}(\mathbf{V}) \quad (39)$$

for

$$\begin{aligned}\mathcal{Q}^{j,(\kappa)}_{i,p(j)}(\mathbf{V}) \triangleq\; & \mathbf{V}^H_{i,p(j)}\mathbf{H}^H_{i,i,j}\mathcal{M}^{p(j)}_{i,j}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}^{(\kappa)}_{i,p(j)} \\ & + (\mathbf{V}^{(\kappa)}_{i,p(j)})^H\mathbf{H}^H_{i,i,j}\mathcal{M}^{p(j)}_{i,j}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)} \\ & - (\mathbf{V}^{(\kappa)}_{i,p(j)})^H\mathbf{H}^H_{i,i,j}\mathcal{M}^{p(j)}_{i,j}(\mathbf{V}^{(\kappa)})^{-1}\Big[\mathcal{L}^{p(j),(\kappa)}_{i,j}(\mathbf{V}) \\ & + \sigma^2\mathbf{I}_{N_r}\Big]\mathcal{M}^{p(j)}_{i,j}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}^{(\kappa)}_{i,p(j)}. \end{aligned} \quad (40)$$

Therefore,

$$\begin{aligned}&\mathcal{R}^j_{i,p(j)}(\mathbf{V}) \\ &= \ln\Big|\mathbf{I}_L + (\mathbf{V}_{i,p(j)})^H\mathbf{H}^H_{i,i,j}\mathcal{M}^{p(j)}_{i,j}(\mathbf{V})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)}\Big| \\ &\geq \ln\Big|\mathbf{I}_L + \widetilde{\mathcal{Q}}^{j,(\kappa)}_{i,p(j)}(\mathbf{V})\Big| \quad (41) \\ &\geq \mathcal{R}^j_{i,p(j)}(\mathbf{V}^{(\kappa)}) + L \\ &\quad - \Big\langle\mathbf{I}_L + \widetilde{\mathcal{Q}}^{j,(\kappa)}_{i,p(j)}(\mathbf{V}^{(\kappa)}), \Big(\mathbf{I}_L + \widetilde{\mathcal{Q}}^{j,(\kappa)}_{i,p(j)}(\mathbf{V})\Big)^{-1}\Big\rangle \quad (42) \\ &\geq \widetilde{\mathcal{R}}^{j,(\kappa)}_{i,p(j)}(\mathbf{V}) \quad (43) \end{aligned}$$

for

$$\begin{aligned}\widetilde{\mathcal{R}}^{j,(\kappa)}_{i,p(j)}(\mathbf{V}) \triangleq\; & \mathcal{R}^j_{i,p(j)}(\mathbf{V}^{(\kappa)}) + L \\ & - \Big\langle\mathbf{I}_L + (\mathbf{V}^{(\kappa)}_{i,p(j)})^H\mathbf{H}^H_{i,i,j}\mathcal{M}^{p(j)}_{i,j}(\mathbf{V}^{(\kappa)})^{-1} \\ & \times\mathbf{H}_{i,i,j}\mathbf{V}^{(\kappa)}_{i,p(j)}, \Big(\mathbf{I}_L + \mathcal{Q}^{j,(\kappa)}_{i,p(j)}(\mathbf{V})\Big)^{-1}\Big\rangle, \end{aligned} \quad (44)$$

which is a concave function. Inequality (41) follows from (33) and the condition

$$\ln|\mathbf{I}_L+\mathbf{X}|\geq\ln|\mathbf{I}_L+\mathbf{Y}|,\ \forall\ \mathbf{X}\succeq\mathbf{Y}\succeq\mathbf{0}.$$

Inequality (42) follows by applying (32) and using equality

$$\ln\big|\mathbf{I}_L+\widetilde{\mathcal{Q}}_{i,p(j)}^{j,(\kappa)}(\mathbf{V}^{(\kappa)})\big|=\mathcal{R}_{i,p(j)}^{j}(\mathbf{V}^{(\kappa)}).$$

Inequality (43) follows from inequality (33), equality (35), and the condition

$$\Big\langle\mathbf{M},\mathbf{X}^{-1}\Big\rangle\geq\Big\langle\mathbf{M},\mathbf{Y}^{-1}\Big\rangle,\ \forall\ \mathbf{M}\succeq\mathbf{0},\mathbf{Y}\succeq\mathbf{X}\succ\mathbf{0}.$$

Analogously, the rate function $\mathcal{R}_{i,p(j)}^{p(j)}(\mathbf{V})$ and throughput function $\mathcal{R}_{i,j}(\mathbf{V})$ defined from (5) and (10) are lower bounded by

$$\mathcal{R}_{i,p(j)}^{p(j)}(\mathbf{V})\geq\widetilde{\mathcal{R}}_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V})\tag{45}$$

over the trust region

$$\mathcal{L}_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V})\succeq\mathbf{0},\tag{46}$$

and

$$\mathcal{R}_{i,j}(\mathbf{V})\geq\widetilde{\mathcal{R}}_{i,j}^{(\kappa)}(\mathbf{V})\tag{47}$$

over the trust region

$$\mathcal{L}_{i,j}^{p,(\kappa)}(\mathbf{V})\succeq\mathbf{0},\tag{48}$$

for linear mappings

$$\mathcal{L}_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V})\triangleq\sum_{(s,l)\in\mathcal{S}\setminus\{(i,p(j))\}}\mathbf{H}_{s,i,p(j)}\Big[\mathbf{V}_{s,l}(\mathbf{V}_{s,l}^{(\kappa)})^H+\mathbf{V}_{s,l}^{(\kappa)}\mathbf{V}_{s,l}^H-\mathbf{V}_{s,l}^{(\kappa)}(\mathbf{V}_{s,l}^{(\kappa)})^H\Big]\mathbf{H}_{s,i,p(j)}^H\tag{49}$$

and

$$\mathcal{L}_{i,j}^{p,(\kappa)}(\mathbf{V})\triangleq\sum_{(s,l)\in\mathcal{S}\setminus\{(i,p(j)),(i,j)\}}\mathbf{H}_{s,i,j}\Big[\mathbf{V}_{s,l}(\mathbf{V}_{s,l}^{(\kappa)})^H+\mathbf{V}_{s,l}^{(\kappa)}\mathbf{V}_{s,l}^H-\mathbf{V}_{s,l}^{(\kappa)}(\mathbf{V}_{s,l}^{(\kappa)})^H\Big]\mathbf{H}_{s,i,j}^H,\tag{50}$$

and for concave functions

$$\widetilde{\mathcal{R}}_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V})\triangleq\mathcal{R}_{i,p(j)}^{p(j)}(\mathbf{V}^{(\kappa)})+L-\Big\langle\mathbf{I}_L+(\mathbf{V}_{i,p(j)}^{(\kappa)})^H\mathbf{H}_{i,i,p(j)}^H\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1}\times\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}^{(\kappa)},\Big(\mathbf{I}_L+\mathcal{Q}_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V})\Big)^{-1}\Big\rangle\tag{51}$$

and

$$\widetilde{\mathcal{R}}_{i,j}^{(\kappa)}(\mathbf{V})\triangleq\mathcal{R}_{i,j}^{(\kappa)}(\mathbf{V}^{(\kappa)})+L-\Big\langle\mathbf{I}_L+(\mathbf{V}_{i,j}^{(\kappa)})^H\mathbf{H}_{i,i,j}^H\mathcal{M}_{i,j}^{p}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}^{(\kappa)},\Big(\mathbf{I}_L+\mathcal{Q}_{i,j}^{(\kappa)}(\mathbf{V})\Big)^{-1}\Big\rangle,\tag{52}$$

Algorithm 2 SDP-based path-following algorithm for the STM (12) in MIMO-NOMA

Initialization: Set $\kappa:=0$ and solve (57) to generate an initial feasible point $\mathbf{V}^{(0)}$ for constraints (12b)-(12d).

1: **repeat**
2: Solve the semi-definite program (56) to obtain the optimal solution: $\mathbf{V}^\star$.
3: Update $\mathbf{V}^{(\kappa+1)}:=\mathbf{V}^\star$.
4: Set $\kappa:=\kappa+1$.
5: **until** Convergence

with

$$\begin{aligned}\mathcal{Q}_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V})\triangleq\ &(\mathbf{V}_{i,p(j)}^{(\kappa)})^H\mathbf{H}_{i,i,p(j)}^H\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}\\&+(\mathbf{V}_{i,p(j)})^H\mathbf{H}_{i,i,p(j)}^H\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}^{(\kappa)}\\&-(\mathbf{V}_{i,p(j)}^{(\kappa)})^H\mathbf{H}_{i,i,p(j)}^H\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1}\Big[\mathcal{L}_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V})\\&+\sigma^2\mathbf{I}_{N_r}\Big]\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}^{(\kappa)}\end{aligned}\tag{53}$$

and

$$\begin{aligned}\mathcal{Q}_{i,j}^{(\kappa)}(\mathbf{V})\triangleq\ &(\mathbf{V}_{i,j}^{(\kappa)})^H\mathbf{H}_{i,i,j}^H\mathcal{M}_{i,j}^{p}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\\&+\ (\mathbf{V}_{i,j})^H\mathbf{H}_{i,i,j}^H\mathcal{M}_{i,j}^{p}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}^{(\kappa)}\\&-\ (\mathbf{V}_{i,j}^{(\kappa)})^H\mathbf{H}_{i,i,j}^H\mathcal{M}_{i,j}^{p}(\mathbf{V}^{(\kappa)})^{-1}\Big[\mathcal{L}_{i,j}^{p,(\kappa)}(\mathbf{V})\\&+\ \sigma^2\mathbf{I}_{N_r}\Big]\mathcal{M}_{i,j}^{p}(\mathbf{V}^{(\kappa)})^{-1}\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}^{(\kappa)}.\end{aligned}\tag{54}$$

We also define

$$\widetilde{\mathcal{R}}_{i,p(j)}^{(\kappa)}(\mathbf{V})\triangleq\min\Big\{\widetilde{\mathcal{R}}_{i,p(j)}^{j,(\kappa)}(\mathbf{V}),\widetilde{\mathcal{R}}_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V})\Big\}.\tag{55}$$

In summary, at the κ-th iteration, the following SDP, which is an inner approximation of (12), is solved to generate the next feasible point $\mathbf{V}^{(\kappa+1)}$:

$$\max_{\mathbf{V}}\ \widetilde{\mathcal{P}}^{(\kappa)}(\mathbf{V})\triangleq\sum_{i=1}^{N}\sum_{j=1}^{K}\Big(\widetilde{\mathcal{R}}_{i,j}^{(\kappa)}(\mathbf{V})+\widetilde{\mathcal{R}}_{i,p(j)}^{(\kappa)}(\mathbf{V})\Big)\tag{56a}$$

$$\text{subject to }\widetilde{\mathcal{R}}_{i,j}^{(\kappa)}(\mathbf{V})\geq r_{i,j},\ \forall i\in\mathcal{I},\ \forall j\in\mathcal{K}_1,\tag{56b}$$

$$\widetilde{\mathcal{R}}_{i,p(j)}^{(\kappa)}(\mathbf{V})\geq r_{i,p(j)},\ \forall i\in\mathcal{I},\ \forall j\in\mathcal{K}_2,\tag{56c}$$

$$(12\text{d}),(37),(46),(48).\tag{56d}$$

The proposed Algorithm 2 generates a sequence $\{\mathbf{V}^{(\kappa)}\}$ of improved points of (12), which also converges to a KKT point.

A feasible point $\mathbf{V}^{(0)}$ for the constraints (12b)-(12d) to initialize Algorithm 2 can be found by invoking the iterations

$$\max_{\mathbf{V}}\ \min_{(i,j)\in\mathcal{S}}\ \min\left\{\frac{\widetilde{\mathcal{R}}_{i,j}^{(\kappa)}(\mathbf{V})}{r_{i,j}},\frac{\widetilde{\mathcal{R}}_{i,p(j)}^{(\kappa)}(\mathbf{V})}{r_{i,p(j)}}\right\}\tag{57a}$$

$$\text{subject to}\quad(12\text{d}),(37),(46),(48)\tag{57b}$$

to reach a value more than or equal to 1 in satisfying (12b)-(12d).

Complexity analysis: The SDP (56) involves $N(1+3K)$ quadratic constraints, $3NK$ semi-definite constraints with N_r rows and $n = KN(2N_tL+1)$ real decision variables. For $m_{\text{SDP}} \triangleq N(1+3K)+3NKN_r$, its computational complexity is $\mathcal{O}(n^2 m_{\text{SDP}}^{2.5} + m_{\text{SDP}}^{3.5})$, which is seen higher than that of the convex quadratic problem (28).

V. Tailored algorithm for MISO-NOMA

In this case, all channels are row vectors ($\mathbf{H}_{s,i,j} \in \mathbb{C}^{1\times N_t}$) and $s_{i,j} \in \mathbb{C}$ ($L=1$).

As observed first time in [25], for

$$\bar{\mathbf{V}}_{i,j} = e^{-\jmath \arg(\mathbf{H}_{i,i,j}\mathbf{V}_{i,j})}\mathbf{V}_{i,j} \tag{58}$$

one has $|\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}| = \mathbf{H}_{i,i,j}\bar{\mathbf{V}}_{i,j} = \Re\{\mathbf{H}_{i,i,j}\bar{\mathbf{V}}_{i,j}\} \geq 0$ and $|\mathbf{H}_{i',i,j'}\mathbf{V}_{i,j}| = |\mathbf{H}_{i',i,j'}\bar{\mathbf{V}}_{i,j}|$ for $(i',j') \neq (i,j)$. Therefore, without loss of generality we can replace

$$\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\mathbf{V}_{i,j}^H\mathbf{H}_{i,i,j}^H = \left|\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\right|^2,\ j=1,\ldots,2K$$

by

$$\left(\Re\{\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\}\right)^2,\ j=1,\ldots,2K$$

with

$$\Re\{\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\} \geq 0,\ j=1,\ldots,2K \tag{59}$$

(including $p(j)$ for $j=K+1,...,2K$). Accordingly, write

$$\mathcal{M}_{i,j}(\mathbf{V}) = \sum_{(s,l)\in\mathcal{S}} \left|\mathbf{H}_{s,i,j}\mathbf{V}_{s,l}\right|^2 + \sigma^2.$$

Then the message $\mathbf{s}_{i,p(j)}$ intended for cell-edge UE $(i,p(j))$ is decoded by the cell-center UE (i,j) with the achievable rate

$$r_{i,p(j)}^{j}(\mathbf{V}) = \ln\left(1+\frac{\left|\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)}\right|^2}{\mathcal{M}_{i,j}^{p(j)}(\mathbf{V})}\right), \tag{60}$$

and is decoded by the cell-edge UE $(i,p(j))$ itself with the achievable rate

$$r_{i,p(j)}^{p(j)}(\mathbf{V}) = \ln\left(1+\frac{\left(\Re\{\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}\}\right)^2}{\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V})}\right) \tag{61}$$

where

$$\begin{aligned}\mathcal{M}_{i,j}^{p(j)}(\mathbf{V}) &\triangleq \mathcal{M}_{i,j}(\mathbf{V}) - \left|\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)}\right|^2 \\ &= \sum_{(s,l)\in\mathcal{S}\setminus\{(i,p(j))\}} \left|\mathbf{H}_{s,i,j}\mathbf{V}_{s,l}\right|^2+\sigma^2,\end{aligned}$$

and

$$\begin{aligned}\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V}) &\triangleq \mathcal{M}_{i,p(j)}(\mathbf{V}) - \left|\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}\right|^2 \\ &= \sum_{(s,l)\in\mathcal{S}\setminus\{(i,p(j))\}} \left|\mathbf{H}_{s,i,p(j)}\mathbf{V}_{s,l}\right|^2+\sigma^2.\end{aligned}$$

Also, the message $\mathbf{s}_{i,j}$ intended for the cell-center UE (i,j) is successively decoded by UE (i,j) itself with the throughput

$$r_{i,j}(\mathbf{V}) = \ln\left(1+\frac{\left(\Re\{\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\}\right)^2}{\mathcal{M}_{i,j}^{p}(\mathbf{V})}\right) \tag{62}$$

where

$$\begin{aligned}\mathcal{M}_{i,j}^{p}(\mathbf{V}) &\triangleq \mathcal{M}_{i,j}^{p(j)}(\mathbf{V}) - \left|\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\right|^2 \\ &= \sum_{(s,l)\in\mathcal{S}\setminus\{(i,p(j)),(i,j)\}} \left|\mathbf{H}_{s,i,j}\mathbf{V}_{s,l}\right|^2+\sigma^2.\end{aligned}$$

For $r_{i,p(j)}(\mathbf{V}) \triangleq \min\{r_{i,p(j)}^{j}(\mathbf{V}), r_{i,p(j)}^{p(j)}(\mathbf{V})\}$, the problem (12) in this case is

$$\underset{\mathbf{V}}{\text{maximize}}\ \mathcal{P}(\mathbf{V}) \triangleq \sum_{i=1}^{N}\sum_{j=1}^{K}\Big(r_{i,j}(\mathbf{V}) + r_{i,p(j)}(\mathbf{V})\Big) \tag{63a}$$

$$\text{subject to } r_{i,j}(\mathbf{V}) \geq r_{i,j},\ \forall i\in\mathcal{I},\ \forall j\in\mathcal{K}_1, \tag{63b}$$

$$r_{i,p(j)}^{p(j)}(\mathbf{V}) \geq r_{i,p(j)},\ \forall i\in\mathcal{I},\ \forall j\in\mathcal{K}_2, \tag{63c}$$

$$r_{i,p(j)}^{j}(\mathbf{V}) \geq r_{i,p(j)},\ \forall i\in\mathcal{I},\ \forall j\in\mathcal{K}_2, \tag{63d}$$

$$\sum_{j\in\mathcal{J}} \|\mathbf{V}_{i,j}\|^2 \leq P_i^{\max}, \forall i\in\mathcal{I}. \tag{63e}$$

Due to the above transforms (61) and (62) under condition (59), the nonconvex constraints (63b) and (63c) are expressed by the second-order cone (SOC) constraints

$$\begin{aligned}\Re\big\{\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\big\} &\geq \sqrt{e^{r_{i,j}}-1}\sqrt{\mathcal{M}_{i,j}^{p}(\mathbf{V})}, \\ &\quad \forall i\in\mathcal{I},\ \forall j\in\mathcal{K}_1,\end{aligned} \tag{64}$$

$$\begin{aligned}\Re\big\{\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}\big\} &\geq \sqrt{e^{r_{i,p(j)}}-1}\sqrt{\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V})}, \\ &\quad \forall i\in\mathcal{I},\ \forall j\in\mathcal{K}_2\end{aligned} \tag{65}$$

but the constraint (63d) remains to be nonconvex.

To approximate functions in (63a) we use the inequality

$$\ln(1+z) \geq a(\bar{z}) - b(\bar{z})\frac{1}{z},\quad \forall\ z>0, \bar{z}>0 \tag{66}$$

with

$$0 < a(\bar{z}) \triangleq \ln(1+\bar{z}) + \frac{\bar{z}}{\bar{z}+1},\ 0 < b(\bar{z}) \triangleq \frac{\bar{z}^2}{\bar{z}+1}, \tag{67}$$

whose proof is provided by Appendix C.

Applying (66) for $\bar{z} = z_{i,p(j)}^{j,(\kappa)} \triangleq |\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)}^{(\kappa)}|^2/\mathcal{M}_{i,j}^{p(j)}(\mathbf{V}^{(\kappa)})$ and $z = |\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)}|^2/\mathcal{M}_{i,j}^{p(j)}(\mathbf{V})$ yields

$$\begin{aligned}r_{i,p(j)}^{j}(\mathbf{V}) &\geq a\big(z_{i,p(j)}^{j,(\kappa)}\big) - b\big(z_{i,p(j)}^{j,(\kappa)}\big)\frac{\mathcal{M}_{i,j}^{p(j)}(\mathbf{V})}{\left|\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)}\right|^2} \\ &\geq r_{i,p(j)}^{j,(\kappa)}(\mathbf{V}) \\ &\triangleq a\big(z_{i,p(j)}^{j,(\kappa)}\big) - b\big(z_{i,p(j)}^{j,(\kappa)}\big)\frac{\mathcal{M}_{i,j}^{p(j)}(\mathbf{V})}{\varphi_{i,p(j)}^{j,(\kappa)}(\mathbf{V})}\end{aligned} \tag{68}$$

over the trust region

$$\begin{aligned}\varphi_{i,p(j)}^{j,(\kappa)}(\mathbf{V}) \triangleq\ & 2\Re\big\{\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)}^{(\kappa)}(\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)})^*\big\} \\ & - \left|\mathbf{H}_{i,i,j}\mathbf{V}_{i,p(j)}^{(\kappa)}\right|^2 > 0.\end{aligned} \tag{69}$$

Algorithm 3 Tailored QP-based path-following algorithm for the STM (63) in MISO-NOMA

Initialization: Initialize a feasible point $\mathbf{V}^{(0)}$ for constraints in (63).

κ-th iteration: Solve the convex quadratic program (74) to find the optimal solution $\mathbf{V}^\star$. If $\big|\big(\mathcal{P}(\mathbf{V}^\star)-\mathcal{P}(\mathbf{V}^{(\kappa)})\big)\big/\mathcal{P}(\mathbf{V}^{(\kappa)})\big|\leq\epsilon$, terminate. Otherwise, set $\kappa:=\kappa+1, \mathbf{V}^{(\kappa)}:=\mathbf{V}^\star$ and continue.

Analogously,

$$r_{i,p(j)}^{p(j)}(\mathbf{V}) \geq r_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V}) \triangleq a\big(z_{i,p(j)}^{p(j),(\kappa)}\big)-b\big(z_{i,p(j)}^{p(j),(\kappa)}\big)\frac{\mathcal{M}_{i,p(j)}^{p(j)}(\mathbf{V})}{\varphi_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V})} \quad (70)$$

with

$$\varphi_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V}) \triangleq \Re\big\{\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}^{(\kappa)}\big\}\Big(2\Re\big\{\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}\big\} - \Re\big\{\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}^{(\kappa)}\big\}\Big)$$

over the trust region

$$2\Re\big\{\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}\big\}-\Re\big\{\mathbf{H}_{i,i,p(j)}\mathbf{V}_{i,p(j)}^{(\kappa)}\big\}>0, \quad (71)$$

and

$$r_{i,j}(\mathbf{V}) \geq r_{i,p(j)}^{(\kappa)}(\mathbf{V}) \triangleq a(z_{i,j}^{(\kappa)})-b(z_{i,j}^{(\kappa)})\frac{\mathcal{M}_{i,j}^{p}(\mathbf{V})}{\varphi_{i,j}^{(\kappa)}(\mathbf{V})} \quad (72)$$

with

$$\varphi_{i,j}^{(\kappa)}(\mathbf{V}) \triangleq \Re\big\{\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}^{(\kappa)}\big\}\Big(2\Re\big\{\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\big\} - \Re\big\{\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}^{(\kappa)}\big\}\Big)$$

over the trust region

$$2\Re\big\{\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}\big\}-\Re\big\{\mathbf{H}_{i,i,j}\mathbf{V}_{i,j}^{(\kappa)}\big\}>0. \quad (73)$$

In Algorithm 3, we propose an QP-based path-following algorithm to solve problem (63). At the κ-th iteration it solves the following SOC program to generate the next feasible point $\mathbf{V}^{(\kappa+1)}$:

$$\max_{\mathbf{V}} \quad \mathcal{P}^{(\kappa)}(\mathbf{V}) \triangleq \sum_{i=1}^{N}\sum_{j=1}^{K}\Big(r_{i,j}^{(\kappa)}(\mathbf{V})+r_{i,p(j)}^{(\kappa)}(\mathbf{V})\Big) \quad (74a)$$

$$\text{subject to } r_{i,p(j)}^{j,(\kappa)}(\mathbf{V}) \geq r_{i,p(j)}, \ \forall i\in\mathcal{I}, \ \forall j\in\mathcal{K}_2, \quad (74b)$$

$$(63e), (64), (65), (69), (71), (73) \quad (74c)$$

where $r_{i,p(j)}^{(\kappa)}(\mathbf{V}) \triangleq \min\Big\{r_{i,p(j)}^{j,(\kappa)}(\mathbf{V}), r_{i,p(j)}^{p(j),(\kappa)}(\mathbf{V})\Big\}$.

To find a feasible point for constraints in (63) for initializing Algorithm 3, initialized by a feasible point $\mathbf{V}^{(0)}$ for the convex constraints (63e), (64), and (65), the following SOC based iterations

$$\max_{\mathbf{V}} \min_{(i,j)\in\mathcal{S}_2} \left\{\frac{r_{i,p(j)}^{j,(\kappa)}(\mathbf{V})}{r_{i,p(j)}}\right\} \quad (75a)$$

$$\text{subject to} \quad (63e), (64), (65), (69) \quad (75b)$$

are invoked for reaching a value more or equal to 1 in satisfying constraints in (63).

Complexity analysis: The SOC program (74) involves $m_{\text{SOC}}=N(1+6K)$ quadratic or SOC constraints and $n=KN(2N_tL+1)$ real decision variables. Its computational complexity is $\mathcal{O}(n^2m_{\text{SOC}}^{2.5}+m_{\text{SOC}}^{3.5})$, which is seen higher than that of the convex quadratic problem (28).

VI. Numerical Results

In this section we use numerical examples to evaluate the performance of the proposed algorithms. A system topology shown in Fig. 2 is set up. There are $N=3$ macro cells and 4 UEs per cell with two cell-center UEs and two cell-edge UEs, which are located near to the boundaries with the two adjacent cells. Unless stated otherwise, $N_t=4$ and $N_r=2$ are set for MIMO-NOMA, for which $L=N_r$ is set. Thus, the precoder-matrices $\mathbf{V}_{i,j}$ are of dimension $N_t\times N_r$. The channel matrix between a BS and a UE at a distance d (in kilometres) is generated as $\mathbf{H}=\sqrt{10^{-\sigma_{\mathsf{PL}}/10}}\tilde{\mathbf{H}}$ [26]. Here, σ_{PL} is the path loss (PL) in dB and the entries of $\tilde{\mathbf{H}}$ are independent and identically distributed complex Gaussian variables with zero mean and unit variance. Without loss of generality, the requirement thresholds for all UEs are set as $r_{i,j}=r_{i,p(j)}\equiv\bar{\mathsf{R}}$ and the same power budget $P_i^{\max}=P^{\max}, \forall i\in\mathcal{I}$ is given to all BSs.

For the ease of reference, the other parameters given in Table I including $\bar{\mathsf{R}}$ are used. The error tolerance in the proposed Algorithms is set to $\epsilon=10^{-3}$. The numerical results are obtained using the parser YALMIP [27]. The achieved sum throughput results are divided by $\ln(2)$ to arrive at the unit of bps/channel-use. Each simulation is run 100 times and the result are averaged to arrive at the final figures.

A. Algorithms' Convergence

Fig. 3 shows the typical convergence behavior of the proposed algorithms for a given set of channel realizations that are randomly generated for the two cases. According to Fig. 3(a), both Algorithm 1 and 2 for MIMO-NOMA reach the almost optimal value of sum throughput in 18 and 13 iterations. As expected, Algorithm 2 converges faster than Algorithm 1. On the other hand, according to Fig. 3(b), Algorithm 3 for NOMA-MISO converges very fast reaching the optimal value in 6 iterations.

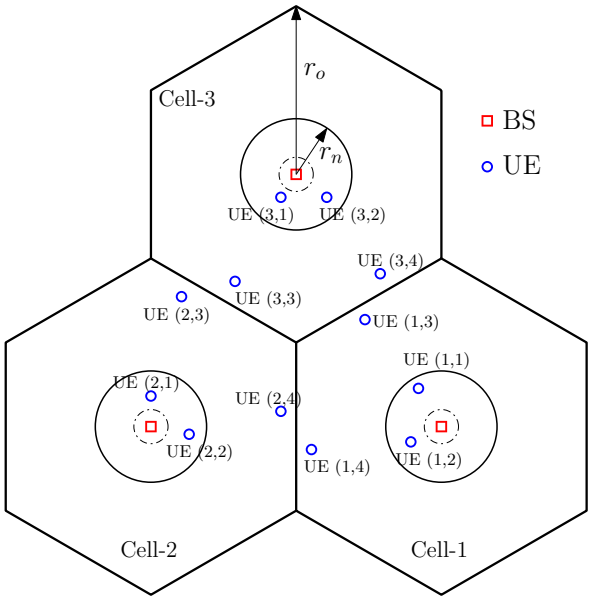
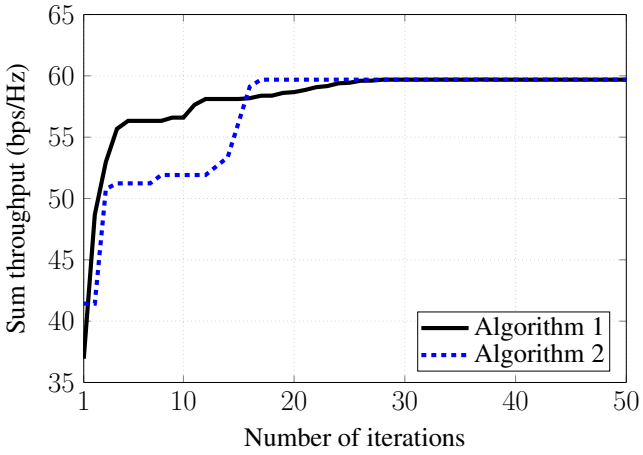


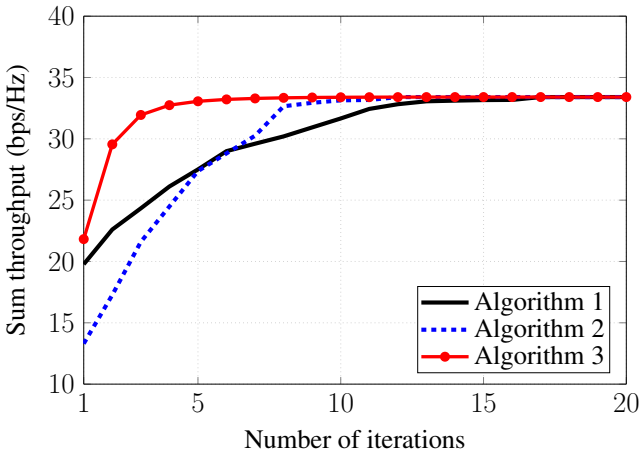
Fig. 2. MIMO-NOMA multi-cell system.

TABLE I
SIMULATION PARAMETERS

Parameters	Value
Carrier frequency/ Bandwidth	2 [GHz]/ 20 [MHz]
Noise power density	-174 [dBm/Hz]
Path loss from the BS to a UE, σ_{PL}	128.1 + 37.6$\log_{10}(d)$ [dB]
Shadowing standard deviation	8 [dB]
Radius of each cell, r_o	500 [m]
Coverage of near UEs, r_n	150 [m]
Distance between BS and nearest UE	$>$ 10 [m]
Threshold $\bar{\mathsf{R}}$	1 [bps/Hz]



(a) MIMO-NOMA networks.



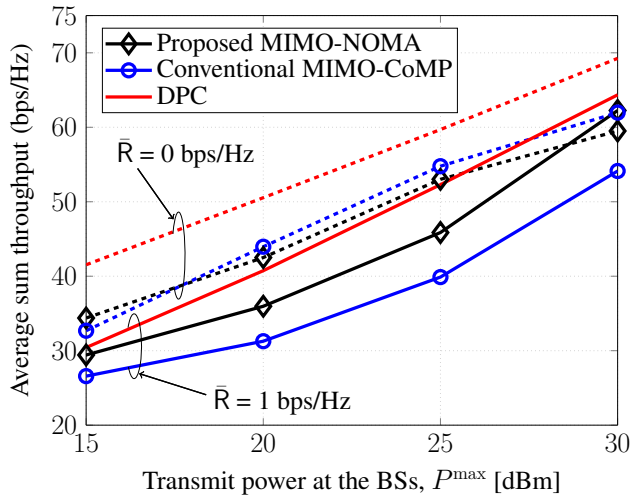
(b) MISO-NOMA networks.

Fig. 3. Convergence pattern for $P^{\max} = 30$ dBm.

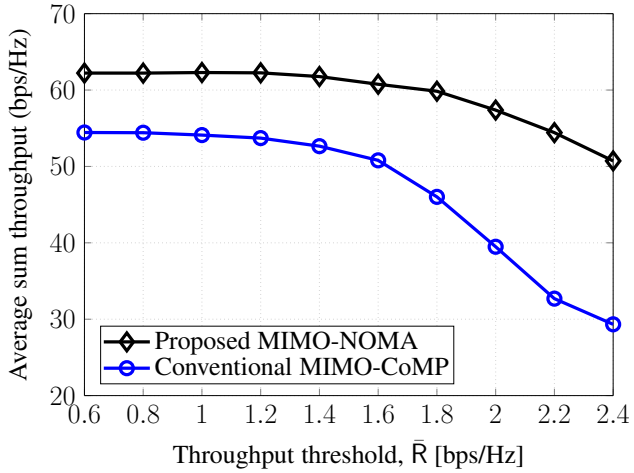
B. Numerical Results for MIMO-NOMA

For the ease of reference, the result achieved by (13) is labeled by "Conventional MIMO-CoMP" whereas that achieved by (12) is labeled by "Proposed MIMO-NOMA." Fig. 4(a) plots the sum throughput versus the power budget $P^{\max}$ under setting $\bar{\mathsf{R}} = \{0, 1\}$ bps/Hz. For $\bar{\mathsf{R}} = 0$ bps/Hz, i.e. there is no UEs' QoS requirement imposed, CoMP slightly outperforms MIMO-NOMA by achieving throughput concentrated at the cell-center UEs of good conditions. Table II details the UEs's throughput distribution for $P^{\max} = 30$ dBm. The high ratio $9.39/0.38 = 24.7$ between the best UE throughput and the worst UE throughput (BWR) implies that CoMP would perform wobbly in the QoS maximization problem (29) (with $r_{i,j} \equiv 1$), which also expresses the system ability to offer the uniform service to UEs. BWR for MIMO-NOMA is $5.1 = 7.13/1.39$ so it is expected to outperform CoMP in maximizing (29). The low throughput 1.46 bps/Hz at UE $(3, 4)$ by DPC is a result of a strong interference from an adjacent cell, which cannot be mitigated by DPC.

For $\bar{\mathsf{R}} = 1$ bps/Hz, MIMO-NOMA of course offers a higher sum throughput than CoMP, where the BSs are seen spending a nearly full power budget $P^{\max}$ in gaining the sum throughput. Increasing $P^{\max}$ also leads to a remarkable gain in sum throughput by NOMA compared with CoMP. The sum throughput by the former also catches up that by DPC. The gain of MIMO-NOMA is a result of canceling interference from intra-cluster interference, as shown in (11). The cell-center UEs in CoMP experience intra-cluster interference that becomes stronger when transmit power increases. The plot of sum throughput versus QoS requirement threshold $\bar{\mathsf{R}} \in [0.6, 2.4]$ bps/Hz is shown by Fig. 4(b) for $P^{\max} = 30$ dBm. The sum throughput are nearly flat for $\bar{\mathsf{R}} \leq 1.2$ bps/Hz and are degraded after that. The BSs in CoMP must allocate much more power to serve cell-edge UEs when QoS threshold increases. As a result, the system sum throughput is dropped quickly. In contrast, the sum throughput of MIMO-NOMA is still slightly sensitive to QoS requirement threshold because



(a) Sum throughput versus $P^{\max}$.



(b) Sum throughput versus $\bar{R}$ with $P^{\max} = 30$ dBm.

Fig. 4. Sum throughput of MIMO-NOMA, (a) versus the transmit power at the BSs and (b) versus the throughput threshold ($N = 3, K = 2, N_t = 4$, and $N_r = 2$).

BSs can tune the power allocation in meeting cell-edge UEs' QoS requirements whenever the cell-center UEs' QoS requirement is easily met.

Fig. 5 shows the impact of the number of UEs per cell and the number of transmit antennas at the BS on the performance of the system. Fig. 5(a) shows that MIMO-NOMA can deliver an acceptable sum throughput for large K. Again, MIMO-NOMA outperforms CoMP in all K. The sum throughput of the both systems decreases from a certain value of K where there is not much degree-of-freedom (DoF) for leveraging multi-user diversity. Interestingly, MIMO-NOMA achieves its best sum throughput for $K = 3$ (6 UEs) while CoMP is peaked at $K = 2$ (4 UEs). Of course, these numbers are not magic and can be changed in other settings. Fig. 5(b) plots the sum throughput vs the number N_t of antennas at the BSs.

Table III details the throughput at UEs under setting

TABLE II
ACHIEVED USER THROUGHPUT (BPS/HZ) IN MULTI-USER MIMO MULTI-CELL SYSTEMS FOR $\bar{R} = 0$ BPS/HZ

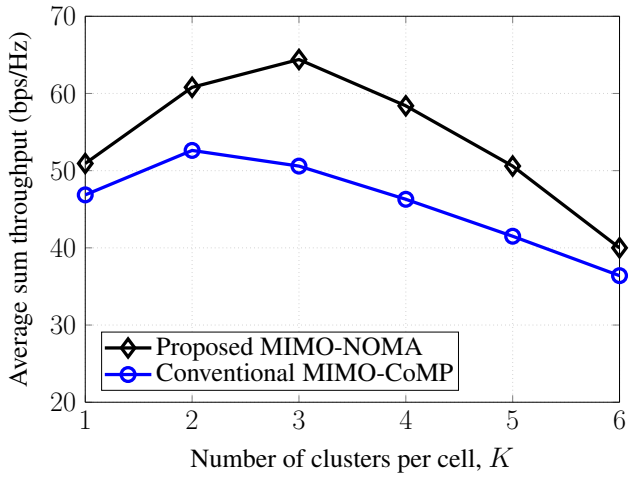
	Throughput per UE			ST per cell		
	NOMA	CoMP	DPC	NOMA	CoMP	DPC
UE (1,1)	6.36	8.27	9.43	18.48	19.11	22.54
UE (1,2)	7.13	9.39	7.86			
UE (1,3)	1.39	1.07	3.27			
UE (1,4)	3.60	0.38	1.98			
UE (2,1)	5.96	9.70	10.76	18.50	23.35	24.01
UE (2,2)	7.02	7.87	8.93			
UE (2,3)	3.59	2.19	2.69			
UE (2,4)	1.93	3.59	1.63			
UE (3,1)	5.04	7.62	10.06	21.27	19.25	22.62
UE (3,2)	7.30	8.44	8.24			
UE (3,3)	4.40	0.63	2.86			
UE (3,4)	4.53	2.56	1.46			
Total ST				**58.25**	**61.71**	**69.17**

TABLE III
ACHIEVED USER THROUGHPUT (BPS/HZ) IN MULTI-USER MIMO MULTI-CELL SYSTEMS

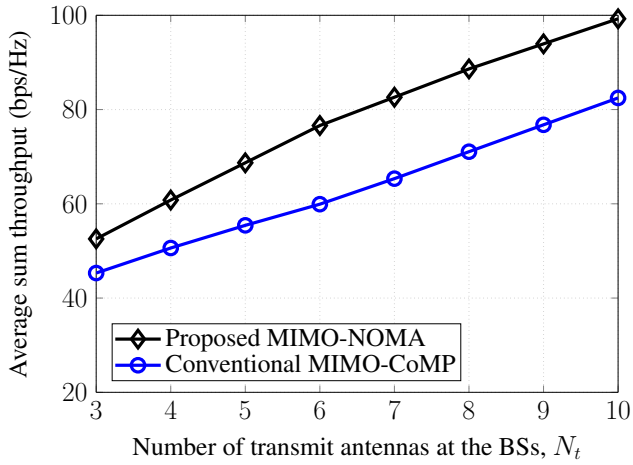
	Throughput per UE		ST per cell		Total ST	
	CoMP	NOMA	CoMP	NOMA	CoMP	NOMA
UE (1,1)	8.37	7.83	20.68	27.16	**59.63**	**77.23**
UE (1,2)	8.74	7.62				
UE (1,3)	2.32	5.12				
UE (1,4)	1.25	6.59				
UE (2,1)	7.81	7.56	18.76	24.34		
UE (2,2)	6.31	6.24				
UE (2,3)	3.13	5.78				
UE (2,4)	1.51	4.76				
UE (3,1)	9.01	8.86	20.19	25.73		
UE (3,2)	6.25	6.92				
UE (3,3)	1.89	6.33				
UE (3,4)	3.04	3.62				

$N_t = 6$ and $P^{\max} = 30$ dBm. Under the same QoS requirement threshold for UEs, the CoMP's throughput is mostly contributed by the cell-center UEs, i.e. CoMP still tends to punish the cell-edge UEs, who are in poor channel condition. Raising the QoS requirement for the cell-edge UEs to counter this discrimination would lead to the risk of CoMP service feasibility. In contrast, in maximizing the system's throughput, MIMO-NOMA offers much fairer and balanced services without contrasting the QoS requirement thresholds so it is very suitable for new quality-of-experience (QoE) services for cell-edge UEs.

By Fig. 6, the system's throughput achieved by the

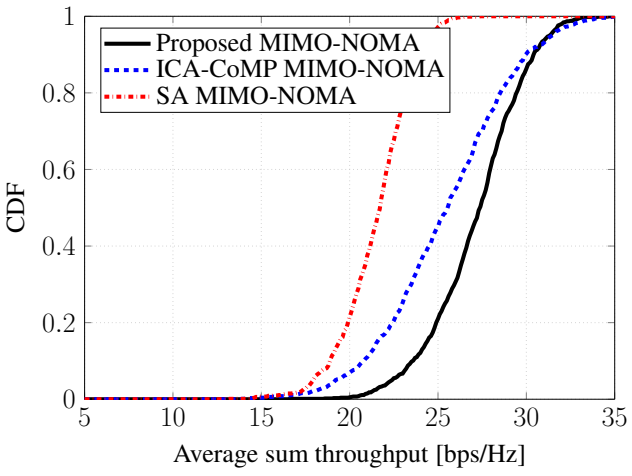


(a) Sum throughput versus K with $N_t = 4$.

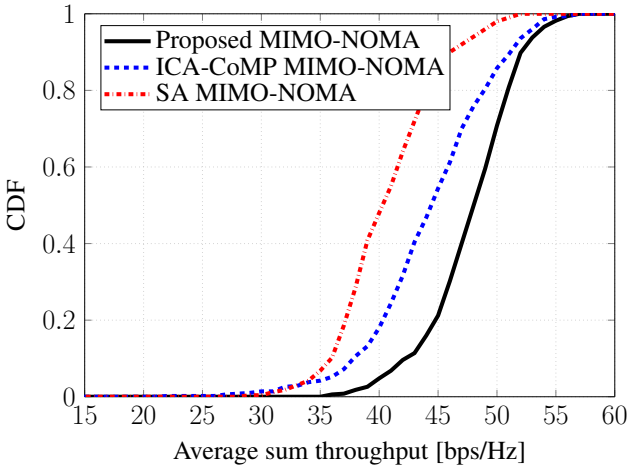


(b) Sum throughput versus N_t with $K = 2$.

Fig. 5. Sum throughput of the MIMO-NOMA, (a) versus the number of clusters per cell and (b) versus the number of transit antennas at the BS ($N = 3, N_r = 2$, and $P^{\max} = 30$ dBm).



(a) CDF versus sum throughput for $P^{\max} = 15$ dBm.



(b) CDF versus sum throughput for $P^{\max} = 30$ dBm.

Fig. 6. CDF of the sum throughput (a) for $P^{\max} = 30$ dBm and (b) for $P^{\max} = 15$ dBm with $N = 2, K = 4, N_t = 5$, and $N_r = 4$.

proposed precoder design is compared to that achieved by signal alignment MIMO-NOMA (SA MIMO-NOMA) [10] and interfering channel alignment CoMP MIMO-NOMA (ICA-CoMP MIMO-NOMA) [11] in the two-cell scenario with $K = 4$ users, $N_t = 5$, while $N_r = 4$ is set to make the signal and channel alignment feasible. In SA MIMO-NOMA, the inter-cluster interference is canceled by detection vectors based SA technique at the UEs and zero-forcing (ZF) based precoder matrix at the BSs. In ICA-CoMP MIMO-NOMA, a receive beamformer is constructed at the cell-edge UEs to align the interfering channels, and then a transmit beamformer based on the null space at the BS is designed to ensure zero inter-cell and inter-cluster interference. For the cell-center UEs, a ZF decoder is designed to cancel the inter-cluster interference only. To ensure a fair comparison, it is additionally assumed that the cell-center UEs do not experience inter-cell interference as they are far away from the neighboring cell in practice [11]. The two different settings of $P^{\max}$ = (15 dBm, 30 dBm), according to 3GPP TR 36.942 v.9.0.1 with a 46-dBm maximum transmit power for the 20 MHz bandwidth are under consideration. Fig. 6 plots the cumulative distribution function (CDF) of the sum throughput. As expected, MIMO-NOMA outperforms both ICA-CoMP MIMO-NOMA and SA MIMO-NOMA. Specifically, by controlling the interference to the cell-edge UEs more efficiently, MIMO-NOMA reaches 1.5 bps/Hz and 5.9 bps/Hz higher than the ICA-CoMP MIMO-NOMA and SA MIMO-NOMA, respectively, in about 60% of the simulated trials with $P^{\max}$ = 15 dBm (see Fig. 6(a)). With $P^{\max}$ = 30 dBm, it is even more essential (see Fig. 6(b)).

C. Numerical Results for MISO-NOMA

The performance of MISO-NOMA achieved by (63) is compared to that achieved by the two-stage beamforming in [12]. The key idea of the two-stage beamforming is that

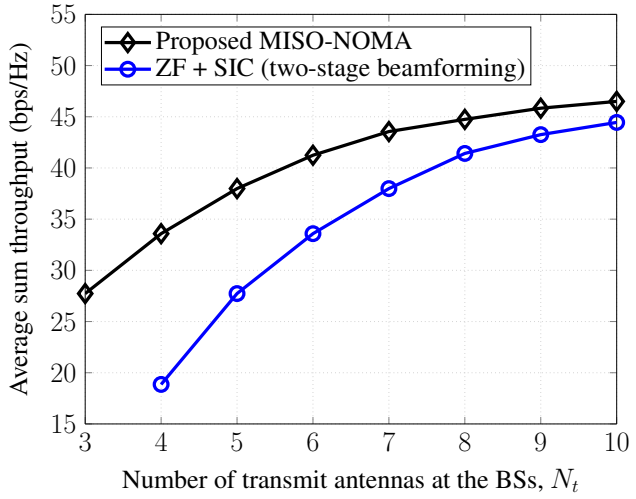


Fig. 7. Average sum throughput of the MISO-NOMA versus the number of transit antennas at the BS ($N = 3, K = 2$, and $P^{\max} = 30$ dBm).

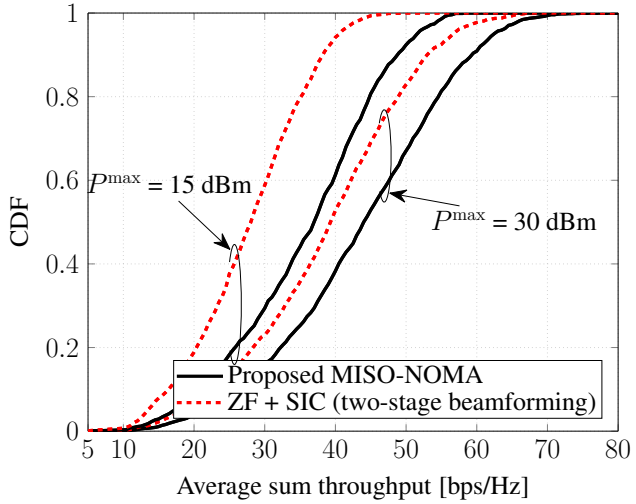
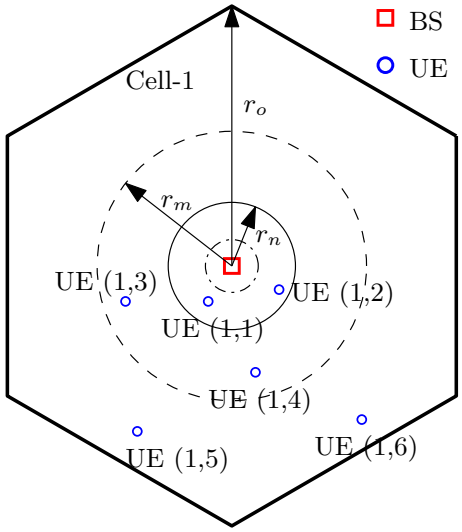


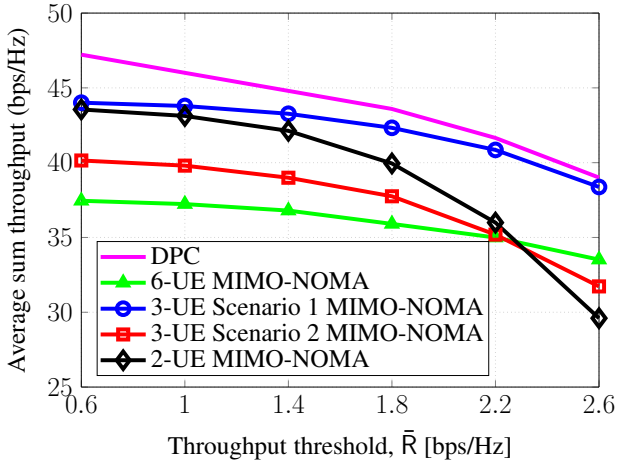
Fig. 8. CDF of the sum throughput for two different settings of the transmit power ($N = 3, K = 2, N_t = 6$, and $N_r = 1$).

ZF beamforming is employed at the BS first to cancel the inter-pair interference and then SIC is used for each pair of UEs. The optimal solution for two-stage beamforming can be easily found by using Algorithm 3.

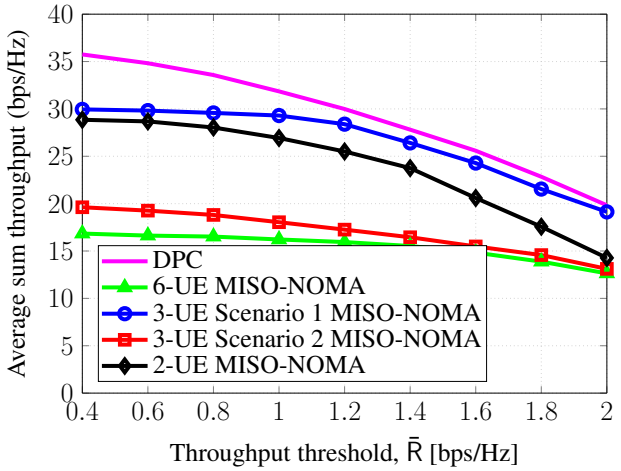
Fig. 7 plots the total sum throughput vs. the number of transmit antennas at the BS under setting $N = 3$, $K = 2$ and $P^{\max} = 30$ dBm. Note that two-stage beamforming requires $N_t \geq 2(K-1)+2 = 4$. Unsurprisingly, MISO-NOMA achieves a better sum throughput than the two-stage beamforming. Two-stage beamforming achieves closer performance to MISO-NOMA as the number N_t of transmit antennas increases, providing more degrees of freedom to leverage multi-user diversity. Fig. 8 plots the CDF of the sum throughput at $N_t = 6$. The performance gap is narrower at a higher power budget.



(a) Simulation setup considered in Fig. 9(b) and Fig. 9(c) with $r_n = 100$ m, $r_m = 250$ m, and $r_o = 500$ m.



(b) Average sum throughput of the MIMO-NOMA versus $\bar{R}$ for $N_r = 2$.



(c) Average sum throughput of the MISO-NOMA versus $\bar{R}$ for $N_r = 1$.

Fig. 9. Average sum throughput (b) for the MIMO-NOMA, and (c) for the MISO-NOMA with different cluster sizes ($N_t = 12$ and $P^{\max} = 30$ dBm).

D. Comparison for Different Cluster Sizes

As the last numerical example, we investigate the system performance in a single-cell scenario of more than two UEs grouped to create a virtual cluster for NOMA. There are 6 UEs in total, which are randomly placed in three different areas, as shown by Fig. 9(a). Two cell-center UEs are located inside the disc of radius r_n = 100 m, two cell-middle UEs are located inside the ring of inner radius 100 m and outer radius 250 m, and two cell-edge UEs are located inside the ring of inner radius 250 m and outer radius 500 m. Different cluster sizes are considered: two UEs per cluster, three UEs per cluster, and six UEs per cluster. For two-UE-per-cluster, a cell-center UE is randomly paired with a cell-middle UE, while the other is randomly paired with a cell-edge UE. The unpaired cell-middle UE and cell-edge are then paired to create the third cluster. There are two scenarios in grouping for three-UE-per-cluster.

- scenario-1 (*more distinct channel conditions*): each cluster consists of a cell-center UE, a cell-middle UE and a cell-edge UE;
- scenario-2 (*less distinct channel conditions*): each cluster consists of a cell-center UE.

The order of decoding messages for UEs in the same cluster of size three is as follows: the message for a cell-edge UE is decoded by all UEs, the message for a cell-middle UE is decoded by itself and the third UE by canceling the previously decoded message for the cell-edge UE from the intra-cluster, and the message for the third UE is decoded by itself only by canceling all previously decoded messages from the intra-cluster interference. Analogously, the messages for UEs $(1,1)$, $(1,2)$, $(1,3)$, $(1,4)$, $(1,5)$ and $(1,6)$ are successively decoded and canceled from the intra-cell interference in the case of six-UE-per-cluster, i.e. all UEs of the same cell have NOMA. The proposed algorithms are easily adapted for solution of the corresponding sum throughput maximization problems.

Fig. 9(b) and Fig. 9(c) plot the sum throughput achieved by different clustering versus throughput threshold for MIMO-NOMA and MISO-NOMA under setting $N_t = 12$ and $P^{\max} = 30$ dBm. In general, the sum throughput achieved by the three-UE-per-cluster and six-UE-per-cluster based schemes are dropped less than that achieved by two-UE-per-cluster based one when the threshold $\bar{\mathsf{R}}$ raises. Specifically, in Fig. 9(b), the sum throughput of the three-UE scenario 2 and six-UE are worse than the two-UE for $\bar{\mathsf{R}} < 2.4$ bps/Hz and vice versa. The BS will allocate a much higher transmit power to the UE of the worst channel condition in the three-UE and six-UE schemes than in the two-UE scheme to meet the QoS constraints. In other words, the cell-edge UEs' throughput is significantly improved in larger cluster sizes.

Notably, the sum throughput by three-UE scenario 1 catches up that by DPC for larger $\bar{\mathsf{R}}$ in both Fig. 9(b) and Fig. 9(c). In addition, six-UE scheme cannot provide a good sum throughput since large UEs per cluster may have error propagation in SIC leading to drastically reduce NOMA performance. Another interesting observation is that the sum throughput of the three-UE scenario 1 outperforms that of the other schemes. Recalling that NOMA is more efficient by exploiting their channel condition differences, i.e. 3-UE scenario 1 with more distinct channel conditions. Consequently, a larger cluster size is recommended for more distinct channel conditions while a smaller cluster size is recommended for less distinct channel conditions.

VII. Conclusions

We have addressed the problem of sum throughput maximization in NOMA based systems by proposing new path-following optimization algorithms. Numerical examples with realistic parameters have confirmed their fast convergence to an optimal solution. They reveal that NOMA not only helps increase the cell-edge UEs' throughput substantially but also achieves much higher total sum throughput. The appropriate size of UE cluster with more distinctive channel gains has also been shown to achieve remarkable gains in NOMA systems.

Appendix A
Proof for inequality (31)

By [28, Appendix B]

$$\big(\alpha\mathbf{V}_1+\beta\mathbf{V}_2\big)^H\big(\alpha\mathbf{X}_1+\beta\mathbf{X}_2\big)^{-1}\big(\alpha\mathbf{V}_1+\beta\mathbf{V}_2\big)\preceq \alpha\mathbf{V}_1^H\mathbf{X}_1^{-1}\mathbf{V}_1+\beta\mathbf{V}_2^H\mathbf{X}_2^{-1}\mathbf{V}_2 \quad (76)$$

for all $\alpha\geq 0$, $\beta\geq 0$, $\alpha+\beta=1$ and $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{X}_1\succ\mathbf{0}$, $\mathbf{X}_2\succ\mathbf{0}$. This means for all $\mathbf{x}$, function

$$f(\mathbf{V},\mathbf{X})=\mathbf{x}^H\mathbf{V}^H\mathbf{X}^{-1}\mathbf{V}\mathbf{x} \quad (77)$$

is convex. Then, for all $\mathbf{V}$, $\bar{\mathbf{V}}$, $\mathbf{X}\succ\mathbf{0}$, $\bar{\mathbf{X}}\succ\mathbf{0}$ it is true that [23]

$$\begin{aligned} f(\mathbf{V},\mathbf{X}) &\geq f(\bar{\mathbf{V}},\bar{\mathbf{X}})+\big\langle\nabla f(\bar{\mathbf{V}},\bar{\mathbf{X}}),(\mathbf{V},\mathbf{X})-(\bar{\mathbf{V}},\bar{\mathbf{X}})\big\rangle \\ &=\mathbf{x}^H\Big[\bar{\boldsymbol{V}}^H\bar{\boldsymbol{X}}^{-1}\mathbf{V}+\mathbf{V}^H\bar{\boldsymbol{X}}^{-1}\bar{\boldsymbol{V}} \\ &\qquad -\bar{\boldsymbol{V}}^H\bar{\boldsymbol{X}}^{-1}\mathbf{X}\bar{\boldsymbol{X}}^{-1}\bar{\boldsymbol{V}}\Big]\mathbf{x}, \end{aligned} \quad (78)$$

i.e.,

$$\begin{aligned} \mathbf{x}^H\mathbf{V}^H\mathbf{X}^{-1}\mathbf{V}\mathbf{x}\geq\mathbf{x}^H\Big[&\bar{\boldsymbol{V}}^H\bar{\boldsymbol{X}}^{-1}\mathbf{V}+\mathbf{V}^H\bar{\boldsymbol{X}}^{-1}\bar{\boldsymbol{V}} \\ &-\bar{\boldsymbol{V}}^H\bar{\boldsymbol{X}}^{-1}\mathbf{X}\bar{\boldsymbol{X}}^{-1}\bar{\boldsymbol{V}}\Big]\mathbf{x},\ \forall\mathbf{x} \end{aligned} \quad (79)$$

proving (31).

APPENDIX B
PROOF FOR INEQUALITY (32)

Since function $\ln|\mathbf{X}|$ is concave on $\mathbf{X} \succ \mathbf{0}$, it is true that [23]

$$-\ln|\mathbf{A}| \geq -\ln|\mathbf{B}| - \left\langle \mathbf{B}^{-1}, \mathbf{A}-\mathbf{B} \right\rangle, \quad \forall\ \mathbf{A} \succ \mathbf{0}, \mathbf{B} \succ \mathbf{0} \tag{80}$$

or equivalently

$$\ln|\mathbf{A}^{-1}| \geq \ln|\mathbf{B}^{-1}| - \left\langle \mathbf{B}^{-1}, \mathbf{A}-\mathbf{B} \right\rangle, \quad \forall\ \mathbf{A} \succ \mathbf{0}, \mathbf{B} \succ \mathbf{0}. \tag{81}$$

Then (32) follows by substituting $\mathbf{X} = \mathbf{A}^{-1}$ and $\bar{\mathbf{X}} = \mathbf{B}^{-1}$ into (81).

APPENDIX C
PROOF FOR INEQUALITY (66)

By [29, Th. 6] function $\ln(1+x^{-1})$ is convex on $x > 0$ so for all $x > 0$ and $\bar{x} > 0$, it is true that [23]

$$\ln\left(1+x^{-1}\right) \geq \ln\left(1+\bar{x}^{-1}\right) + \left[(1+\bar{x})^{-1} - \bar{x}^{-1}\right](x-\bar{x}). \tag{82}$$

Inequality (66) then follows by substituting $z = x^{-1}$ and $\bar{z} = \bar{x}^{-1}$ into (82).

REFERENCES

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[2] J. Lee, Y. Kim, H. Lee, B. L. Ng, D. Mazzarese, J. Liu, W. Xiao, and Y. Zhou, "Coordinated multipoint transmission and reception in LTE-Advanced systems," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 44–50, Nov. 2012.

[3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[4] H. Weingarten, Y. Steinberg, and S. S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[5] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C.-L. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[6] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, June 2013, pp. 1–5.

[7] A. Benjebbour, A. Li, Y. Kishiyama, H. Jiang, and T. Nakamura, "System-level performance of downlink NOMA combined with SU-MIMO for future LTE enhancements," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2014, pp. 706–710.

[8] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[9] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.

[10] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4483–4454, June 2016.

[11] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, "Coordinated beamforming for multi-cell MIMO-NOMA," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 84–87, 2016.

[12] J. Choi, "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 791–800, Mar. 2015.

[13] Z. Chen, Z. Ding, P. Xu, and X. Dai, "Optimal precoding for a QoS optimization problem in two-user MISO-NOMA downlink," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1263–1266, June 2016.

[14] Z. Chen, Z. Ding, X. Dai, and G. K. Karagiannidis, "On the application of quasi-degradation to MISO-NOMA downlink," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6174–6189, Dec. 2016.

[15] Z. Chen, Z. Ding, and X. Dai, "Beamforming for combating inter-cluster and intra-cluster interference in hybrid NOMA systems," *IEEE Access*, vol. 4, pp. 4452–4463, 2016.

[16] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 76–88, Jan. 2016.

[17] Q. Sun, S. Han, C. L. I, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 405–408, Aug. 2015.

[18] J. Choi, "On the power allocation for MIMO-NOMA systems with layered transmissions," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3226–3237, May 2016.

[19] Y. Liu, M. Elkashlan, Z. Ding, and G. K. Karagiannidis, "Fairness of user clustering in MIMO non-orthogonal multiple access systems," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1465–1468, July 2016.

[20] H. H. M. Tam, H. D. Tuan, and D. T. Ngo, "Successive convex quadratic programming for quality-of-service management in full-duplex MU-MIMO multicell networks," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2340–2353, June 2016.

[21] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[22] E. Che, H. D. Tuan, H. H. M. Tam, and H. H. Nguyen, "Successive interference mitigation in multiuser MIMO interference channels," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2185–2199, June 2015.

[23] H. Tuy, *Convex Analysis and Global Optimization (second edition)*. Springer, 2016.

[24] H. H. Kha, H. D. Tuan, and H. H. Nguyen, "Fast global optimal power allocation in wireless networks by local D.C. programming," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 510–515, Feb. 2012.

[25] A. Wiesel, Y. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[26] "Radio Frequency (RF) requirements for LTE Pico Node B," Oct. 2012, ETSI Technical Report 136.931.

[27] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. 2004 IEEE Inter. Symposium on Computer Aided Control Systems Design*, Sept. 2004, pp. 284–289.

[28] U. Rashid, H. D. Tuan, H. H. Kha, and H. H. Nguyen, "Joint optimization of source precoding and relay beamforming in wireless MIMO relay networks," *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 488–499, Feb. 2014.

[29] H. M. T. Ho, H. D. Tuan, D. T. Ngo, T. Q. Duong, and H. V. Poor, "Joint load balancing and interference management for small-cell heterogeneous networks with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 872–884, Feb. 2017.

Van-Dinh Nguyen (S'14) received his B.S. degree in Telecommunications from Ho Chi Minh City University of Technology (Bach Khoa University), Vietnam, in 2012, and his M.S. degree in Electronic Engineering from Soongsil University, Seoul, Korea, in 2015. He was a visiting student at Queen's University Belfast, UK (June-July 2015 and August 2016). Currently, he is pursuing his Ph.D. degree in Electronic Engineering at Soongsil University, Seoul, Korea. His research interests include wireless communications, information theory, physical layer security, energy harvesting, cognitive radio, and nonorthogonal multiple access.

Hoang Duong Tuan received the Diploma (Hons.) and Ph.D. degrees in applied mathematics from Odessa State University, Ukraine, in 1987 and 1991, respectively. He spent nine academic years in Japan as an Assistant Professor in the Department of Electronic-Mechanical Engineering, Nagoya University, from 1994 to 1999, and then as an Associate Professor in the Department of Electrical and Computer Engineering, Toyota Technological Institute, Nagoya, from 1999 to 2003. He was a Professor with the School of Electrical Engineering and Telecommunications, University of New South Wales, from 2003 to 2011. He is currently a Professor with the Centre for Health Technologies, University of Technology Sydney. He has been involved in research with the areas of optimization, control, signal processing, wireless communication, and biomedical engineering for more than 20 years.

Trung Q. Duong (S'05–M'12–SM'13) received his Ph.D. degree in Telecommunications Systems from Blekinge Institute of Technology (BTH), Sweden in 2012. Since 2013, he has joined Queen's University Belfast, UK as a Lecturer (Assistant Professor). His current research interests include small-cell networks, physical layer security, energy-harvesting communications, cognitive relay networks. He is the author or co-author of more than 250 technical papers published in scientific journals (138 articles) and presented at international conferences (120 papers).

Dr. Duong currently serves as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IET COMMUNICATIONS, and a Senior Editor for IEEE COMMUNICATIONS LETTERS. He was awarded the Best Paper Award at the IEEE Vehicular Technology Conference (VTC-Spring) in 2013, IEEE International Conference on Communications (ICC) 2014, and IEEE Global Communications Conference (GLOBECOM) 2016. He is the recipient of prestigious Royal Academy of Engineering Research Fellowship (2016-2021).

H. Vincent Poor (S'72–M'77–SM'82–F'87) received the Ph.D. degree in electrical engineering and computer science from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering. During 2006 to 2016, he served as Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, most recently at Berkeley and Cambridge. Dr. Poor's research interests are in the areas of information theory and signal processing, and their applications in wireless networks and related fields. Among his publications in these areas is the recent book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017).

Dr. Poor is a member of the National Academy of Engineering, the National Academy of Sciences, and is a foreign member of the Royal Society. In 1990, he served as President of the IEEE Information Theory Society, and in 2004-07 as the Editor-in-Chief of the IEEE TRANSACTIONS ON INFORMATION THEORY. He received the Marconi and Armstrong Awards of the IEEE Communications Society in 2007 and 2009, respectively. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal, a D.Sc. *honoris causa* from Syracuse University awarded in 2017, and election as a Foreign Member of the National Academy of Engineering of Korea and an Honorary Member of the National Academy of Sciences of Korea, both in 2017.

Oh-Soon Shin (S'00–M'10) received his B.S., M.S., and Ph.D. degrees in Electrical Engineering and Computer Science from Seoul National University, Seoul, Korea, in 1998, 2000, and 2004, respectively. From 2004 to 2005, he was with the Division of Engineering and Applied Sciences, Harvard University, MA, USA, as a Postdoctoral Fellow. From 2006 to 2007, he was a Senior Engineer at Samsung Electronics, Suwon, Korea. In September 2007, he joined the School of Electronic Engineering, Soongsil University, Seoul, Korea, where he is currently an Associate Professor. His research interests include communication theory, wireless communication systems, and signal processing for communication.